THE THEORY OF FUNCTIONS

OF A

REAL VARIABLE

AND

THE THEORY OF FOURIER'S SERIES

CAMBRIDGE UNIVERSITY PRESS WAREHOUSE,
C. F. CLAY, MANAGER.
London: FETTER LANE, E.C.
Glasgow: 50, WELLINGTON STREET.

Leipzig: F. A. BROCKHAUS.
New York: G. P. PUTNAM'S SONS.
Bombay and Calcutta: MACMILLAN & CO., LTD.

'HE THEORY OF FUNCTIONS

OF A

REAL VARIABLE

AND

IE THEORY OF FOURIER'S SERIES

by

E. W. HOBSON, Sc.D., F.R.S.

Fellow of Christ's College, and Stokes Lecturer in Mathematics
in the University of Cambridge

Cambridge
at the University Press

1907

Cambridge:
PRINTED BY JOHN CLAY, M.A.
AT THE UNIVERSITY PRESS.

PREFACE.

THE theory of functions of a real variable, as developed during the last few decades, is a body of doctrine resting, first upon a definite conception of the arithmetic continuum which forms the field of the variable, and which includes a precise arithmetic theory of the nature of a limit, and secondly, upon a definite conception of the nature of the functional relation. The procedure of the theory consists largely in the development, based upon precise definitions, of a classification of functions, according as they possess, or do not possess, certain peculiarities, such as continuity, differentiability, &c., throughout the domain of the variable, or at points forming a selected set contained in that domain. The detailed consequences of the presence, or of the absence, of such peculiarities are then traced out, and are applied for the purpose of obtaining conditions for the validity of the processes of Mathematical Analysis. These processes, which have been long employed in the so-called Infinitesimal Calculus, consist essentially in the ascertainment of the existence, and in the evaluation, of limits, and are subject, in every case, to restrictive assumptions which are necessary conditions of their validity. The object to be attained by the theory of functions of a real variable consists then largely in the precise formulation of necessary and sufficient conditions for the validity of the limiting processes of Analysis. A necessary requisite in such formulation is a language descriptive of particular aggregates of values of the variable, in relation to which functions possess definite peculiarities. This language is provided by the Theory of Sets of Points, also known, in its more general aspect, as the Theory of Aggregates, which contains an analysis of the peculiarities of structure and of distribution in the field of the variable which such sets of points may possess. This theory, which had its origin in the exigencies of a critical theory of functions, and has since received wide applications, not only in Pure Analysis, but also in Geometry, must be regarded as an integral part of the subject. A most important part of the theory of functions is the theory of the representation of functions in a prescribed manner, especially by means of series or sequences of functions of prescribed types. Much progress has recently been made in this part of the subject, results having been obtained which have led to a classification of functions in accordance with the modes of representation of which they are capable. The special case of the conditions of representability of functions by means of trigonometrical series was historically the starting-point in which a great part of the modern development of the theory of functions of a real variable had its origin.

The course of study, of which the present treatise is the outcome, followed an order very similar to the historical order in which the subject was developed. Commencing with the study of Fourier's series, in their application to the problems of Mathematical Physics, and provided with a knowledge of the Differential and Integral Calculus, of the traditional kind in which notions of the nature of continuity and of limits founded on an uncritical use of intuitions of space and time are the stock in trade, I was led, by the difficulties connected with the theory of these series, and through an attempt to understand the literature which deals with them, to a study of the theories of real number, due to Cantor and Dedekind, and to that of the theory of sets of points. A study of the foundations of the Integral Calculus, and of the general theory of functions of a real variable formed the natural continuation of the course. The present work has been written with the object of presenting in a connected form, and of thus rendering more easily accessible than hitherto, the chief results which are to be found scattered through a very large number of memoirs, periodicals, and treatises. I have endeavoured, as far as possible, to fill up gaps in the various theories which occur in different parts of the subject. The proofs of theorems have in many cases been simplified, often in accordance with developments of the theory later in date than the original proofs; other theorems have been given in a form more general than that in which they were first discovered. In the literature of the subject, errors are not infrequent, largely owing to the fact that spatial intuition affords an inadequate corrective of the theories involved, and is indeed in some cases almost misleading. Although I have made every endeavour to attain to accuracy both in form and in substance, it is practically certain that the present work will form no exception to the rule of fallibility. Where I have called attention to what I regard as inadequate statements or errors on the part of other writers, I have done so solely for the purpose of directing the attention of students to the points in question, and with full consciousness that, at least in some cases, close examination might shew that what appeared to me to be erroneous was rather due to some misapprehension on my part of the meaning of the writers to whom reference is made. On some points connected with the theory of aggregates, which are at present matters of controversy, I have expressed definite opinions, although I fully recognize that, on such matters, a dogmatic attitude of mind is at the present time wholly out of place, and not unlikely to be avenged when the points concerned are finally settled to the general satisfaction of mathematicians.

Chapter I contains a discussion of Number, and includes a full account of the theories of Real Number, due to Cantor and Dedekind. Whilst an indication has been given of the fundamental notions upon which the conceptions of cardinal and ordinal numbers rest, I have not attempted to reduce these fundamental notions to a minimum of indefinables from which the whole theory might be deduced by means of formal logic. A slight perusal of the extremely extensive literature of the Philosophy of Arithmetic will shew that any such attempt could only have been made by entering upon a prolonged discussion of a philosophical character, wholly unsuited to a treatise of professedly mathematical complexion, and that any views expressed would have had but little prospect of giving general satisfaction to logicians and philosophers. The modern theory of Real Numbers has been the object of much criticism by philosophers and others. It has been represented that the modern extension of the notion of number to the case of irrational numbers is a sophistical attempt to obliterate the fundamental distinction between the discrete and the continuous. I venture to think that such objections consist, in large part at least, of criticisms of the current terminology of the mathematical theories, especially in respect of the extensions of the use of the word "number," and I think it probable that many of these criticisms would not survive a fair examination of the theories themselves apart from the language in which they are expressed. An appropriate terminology, although a matter of convention, is no doubt a very important matter in relation to such fundamental matters, as it is conducive to clearness of thought; but the substance of the theories is of incomparably greater importance than the forms in which they are expressed, and those theories may be found on examination to be essentially sound, even if their terminology be regarded as in some respects defective.

Chapter II contains an exposition of the theory of sets of points, and includes an account of transfinite cardinal and ordinal Arithmetic, of a somewhat simpler and less general character than will be met with in the treatment of the general theory of aggregates, in Chapter III. Students who do not care to embark upon the discussions in Chapter III will find a study of Chapter II amply sufficient to enable them to apply the ideas there developed in the general theory of functions. A slight account only has been given of the properties of plane sets of points. An account of the important recent investigations which had their origin in Jordan's theorem, that a closed curve divides plane space into two regions, would have occupied more space than was at my disposal. This omission will be less felt than might have been the case, were not an excellent account of this subject to be found in Dr W. H. Young's treatise on the theory of sets of points, which has appeared since this portion of the present work was printed.

In Chapter IV, there will be found a discussion of the main properties of functions, in relation to continuity, discontinuity, &c., and investigations of the properties of important classes of functions. Although the treatise is mainly one on functions of a single variable, a considerable amount of space has been devoted to the consideration of functions of two variables, not only on account of the intrinsic importance of that subject, but because no adequate consideration of the properties of functions of a single variable is possible without the use of functions of two variables, as is seen, for example, from the consideration that a function defined by means of a sequence of functions of a single variable is virtually defined as a limit of a function of two variables.

The foundations of the Integral Calculus, as based upon Riemann's definition of a definite integral, and its extensions, are discussed in Chapter V, where an account of the development of the subject from the point of view of Lebesgue's new definition of the definite Integral is also given. In later parts of the book I have introduced extensions of Lebesgue's definition, to the cases of improper integrals, taken over finite or infinite domains, regarded as the limits of sequences of Lebesgue integrals.

Chapter VI is concerned with functions defined as the limits of sequences of functions, and contains an account of the principal properties of functions represented by series, and a discussion of important matters connected with the modes of convergence of series through whole intervals, or in the neighbourhood of particular points. Various matters relating to the processes of the Integral Calculus, which had not been considered in Chapter V, are here dealt with, because their adequate treatment presupposes a knowledge of the theorems relating to the convergence of sequences of functions. An account of the very general results recently obtained by Baire, relating to the representability of functions by means of series, will be found in this Chapter.

Chapter VII is devoted to the theory of Fourier's series. No apology is needed for the selection of this particular mode of representation of functions for full discussion in a treatise on the theory of functions of a real variable, in view of the historical relation of Fourier's series to the development of the general theory. The history of the theory of Fourier's series is exceedingly instructive, not merely from the point of view of the mathematician, but also from that of the epistemologist. I have therefore endeavoured, in my treatment of the subject, to preserve as much of the historical element as was possible in an account which should contain, in a moderate compass, not only indications of the various stages of development of the subject, but also the most recent results that have been obtained. I have made full use of the greater generality which can be introduced into many of the known results by means of the employment of the theory of integration developed by Lebesgue.

In the preparation of the work, the treatises from which I have most largely drawn information are the German edition of Dini's treatise on the subject, Stolz's *Grundzüge der Differential- und Integral-Rechnung*, Schönflies'

Bericht entitled "*Die Entwickelung der Lehre von den Punktmannigfaltigkeiten*," and the various treatises on different parts of the subject by Borel and Lebesgue. I have consulted a very large number of memoirs, articles, notes, and books, far too numerous to be here particularized. In respect to the references given throughout the book, I wish it to be understood that I have made no attempt to settle questions of priority of discovery. The references given are to be regarded solely as indicating sources of information from which I have drawn, or where more detailed information on the various topics is to be found.

I owe a debt of gratitude to my friend Mr J. W. Sharpe, formerly Fellow of Gonville and Caius College, who has read with the greatest care the proofs of about two-thirds of the book. Many points of difficulty I have fully discussed with him; many obscurities of expression have been removed, and many improvements in substance have been made, owing to the care he has bestowed in reading the proofs. I felt it as a great loss when, owing to a temporary failure of health, he was unable to continue his laborious work. To Dr H. F. Baker, F.R.S., Fellow of St John's College, and Cayley Lecturer in Mathematics, who has kindly read some of the earlier proofs, I owe several valuable suggestions. On several points connected with the treatment of Number in Chapter I, I have had the advantage of consulting Dr James Ward, F.B.A., Fellow of Trinity College, and Professor of Mental Philosophy and Logic in the University.

My thanks are due to the officials of the University Press for the readiness with which they have met my views, and for the care which they have bestowed upon the work connected with the printing. I desire especially to express my sense of the value of the excellent work done by the readers of the Press; to their care is due the elimination of many typographical and other blemishes which would otherwise have remained undetected.

E. W. HOBSON.

CHRIST'S COLLEGE, CAMBRIDGE.
May 15, 1907.

CONTENTS.

CHAPTER I.

NUMBER.

SECTIONS		PAGES
1.	Introduction	1—3
2—4.	Ordinal numbers	3—7
5, 6.	Cardinal numbers	7—10
7—9.	The operations on integral numbers	10—12
10—12.	Fractional numbers	12—16
13, 14.	Negative numbers, and the number zero	16—18
15, 16.	Irrational numbers	19—21
17.	Kronecker's scheme of arithmetization	21—22
18—21.	The Dedekind theory of irrational numbers	22—26
22—27.	The Cantor theory of irrational numbers	26—33
28.	Convergent sequences of real numbers	33—35
29—31.	The arithmetical theory of limits	35—38
32, 33.	Equivalence of the definitions of Dedekind and Cantor	38—40
34.	The non-existence of infinitesimals	40—41
35—37.	The theory of indices	41—45
38, 39.	The representation of real numbers	45—49
40, 41.	The continuum of real numbers	49—51
42.	The continuum given by intuition	52
43, 44.	The straight line as a continuum	53—56

CHAPTER II.

THEORY OF SETS OF POINTS.

45.	Introduction	57—58
46, 47.	The upper and lower boundaries of a set of points	58—60
48.	Limiting point of a set of intervals	60
49—51.	The limiting points and the derivatives of a linear set	61—64
52, 53.	The distribution of points of a set in the interval	64—66
54—56.	Enumerable aggregates	66—70
57, 58.	The power of an aggregate	70—72
59.	The arithmetic continuum	72—74
60—62.	Transfinite ordinal numbers	74—79
63.	The transfinite derivatives of a set of points	79—81
64—71.	Sets of intervals	81—90
72—75.	Non-dense closed and perfect sets	90—97
76.	Properties of the derivatives of sets	98

SECTIONS		PAGES
77—84.	The content and the measure of sets of points	98—108
85, 86.	The content of closed sets	108—110
87, 88.	Ascending sequences of closed sets	110—113
89.	Sets of the first and of the second category	114—117
90, 91.	Diminishing sequences of closed sets	117—119
92, 93.	The common points of a system of open sets	119—122
94, 95.	The analysis of sets in general	123—127
96—99.	Inner limiting sets	127—135
100—108.	Non-linear sets of points	135—146
109.	Sets of sequences of integers	146—147

CHAPTER III.

TRANSFINITE NUMBERS AND ORDER-TYPES.

110.	Introduction	148—149
111.	The cardinal number of an aggregate	149—151
112, 113.	The relative order of cardinal numbers	151—152
114, 115.	The addition and multiplication of cardinal numbers	152—153
116.	Cardinal numbers as exponents	153—154
117, 118.	The smallest transfinite cardinal number	154—156
119, 120.	The equivalence theorem	156—159
121.	Division of cardinal numbers by finite numbers	159—162
122.	The order-type of simply ordered aggregates	162—163
123, 124.	The addition and multiplication of order-types	163—164
125, 126.	The structure of simply ordered aggregates	164—167
127—129.	The order-types η, θ, π	167—170
130—133.	Normally ordered aggregates	170—175
134—136.	The theory of ordinal numbers	175—177
137—139.	The ordinal numbers of the second class	177—180
140.	The cardinal number of the second class of ordinals	180—181
141, 142.	The general theory of Aleph-numbers	181—183
143—145.	The arithmetic of ordinal numbers of the second class	183—185
146.	The theory of order-functions	185—186
147—151.	The cardinal number of the continuum	187—194
152—163.	General discussion of the theory	194—212

CHAPTER IV.

FUNCTIONS OF A REAL VARIABLE.

164.	Introduction	213—214
165, 166.	The functional relation	214—219
167, 168.	The upper and lower limits of functions	219—221
169—171.	The continuity of functions	221—225
172.	Continuous functions defined for a continuous interval	225—226
173, 174.	Continuous functions defined at points of a set	226—228
175.	Uniform continuity	229—230
176—179.	The limits of a function at a point	230—233
180—182.	The discontinuities of functions	233—237

SECTIONS		PAGES
183—185.	Semi-continuous functions	237—240
186, 187.	The classification of discontinuous functions	240—243
188—190.	Point-wise discontinuous functions	243—248
191, 192.	Definition of point-wise discontinuous functions by extension	249—252
193—196.	Functions with limited total fluctuation	252—259
197—199.	The maxima, minima, and lines of invariability of continuous functions	259—263
200, 201.	The derivatives of functions	263—266
202—208.	The differential coefficients of continuous functions	266—275
209—213.	The successive differential coefficients of a continuous function	275—281
214, 215.	Oscillating continuous functions	281—284
216—223.	General properties of derivatives	285—294
224.	Functions with one derivative assigned	294—295
225—228.	The construction of continuous functions	296—299
229.	Functions of two or more variables	299—301
230—235.	Functions of two variables	301—311
236, 237.	Partial differential coefficients	311—316
238—240.	Higher partial differential coefficients	316—321
241, 242.	Maxima and minima of a function of two variables	322—324
243—245.	Properties of a function continuous with respect to each variable	324—328
246—249.	The representation of a square on a linear interval	329—336

CHAPTER V.

INTEGRATION.

250.	Introduction	337
251—254.	The definite integrals of limited functions	338—342
255.	Particular cases of integrable functions	343—344
256.	Properties of the definite integral	344—347
257.	Integrable null-functions and equivalent integrals	347—349
258—264.	The fundamental theorem of the Integral Calculus	349—357
265.	Functions which are linear in each interval of a set	357
266—269.	Mean value theorems	358—364
270—272.	Improper integrals	364—369
273, 274.	Absolutely and conditionally convergent integrals	369—371
275—280.	Existence and properties of absolutely convergent improper integrals	371—377
281—284.	Non-absolutely convergent improper integrals	377—386
285.	The fundamental theorem of the Integral Calculus for the case of improper integrals	387—388
286.	Geometrical interpretation of integration	388—390
287—291.	Lebesgue's theory of integration	390—397
292—300.	Integrals with infinite limits	398—407
301, 302.	Integration by parts	407—409
303—309.	Change of the variable in an integral	410—415
310—312.	Double integration	415—421
313—316.	Repeated integrals	421—430
317.	Properties of the double integral	430—431
318—322.	Improper double integrals	432—442
323, 324.	Double integrals over infinite domains	442—445
325—328.	The transformation of double integrals	445—452

CHAPTER VI.

FUNCTIONS DEFINED BY SEQUENCES.

SECTIONS		PAGES
329, 330.	Introduction	453—454
331.	Non-convergent arithmetic series	455—457
332—334.	Absolutely convergent, and conditionally convergent series	457—461
335.	Series of transfinite type	461—463
336—338.	Double sequences and double series	464—469
339.	Functions represented by series	469—470
340, 341.	Uniform convergence of series	470—474
342.	Non-uniform convergence	474—475
343, 344.	The continuity of the sum-function	475—477
345—348.	Tests of uniform convergence	477—482
349.	Relation of the theory with that of functions of two variables	483—485
350.	The distribution of points of non-uniform convergence	485—487
351, 352.	The limits of a sum-function at a point	487—489
353, 354.	The necessary and sufficient conditions for the continuity of the sum-function	489—495
355—359.	The convergence of power-series	495—499
360, 361.	The convergence of the product of two series	500—501
362—367.	Taylor's series	501—512
368.	Maxima and minima of a function of one variable	512—513
369.	Taylor's theorem for functions of two variables	514—515
370—372.	Maxima and minima of functions of two variables	515—522
373—377.	Functions representable by series of continuous functions	522—532
378.	Baire's classification of functions	532—534
379—388.	The integration of series	534—550
389—393.	The fundamental theorem of the Integral Calculus for Lebesgue integrals	550—559
394.	Integration by parts for Lebesgue integrals	559—560
395—398.	The differentiation of series	561—565
399—403.	Repeated improper integrals	566—576
404—407.	Repeated Lebesgue integrals	576—582
408.	Repeated integrals of unlimited functions	582—586
409—411.	Repeated integrals over an infinite domain	586—594
412—414.	The limit of an integral containing a parameter	594—599
415—420.	Differentiation of an integral with respect to a parameter	599—607
421—423.	The condensation of singularities	607—618
424.	Cantor's method of condensation of singularities	618—620
425, 426.	The construction of continuous non-differentiable functions	620—625
427—429.	The construction of a differentiable everywhere-oscillating function	626—634

CHAPTER VII.

TRIGONOMETRICAL SERIES.

430.	Introduction	635
431, 432.	The problem of vibrating strings	636—638
433.	Special cases of trigonometrical series	638—639
434, 435.	Later history of the theory	639—641

SECTIONS		PAGES
36—440.	The formal expression of Fourier's series	641—648
41—443.	Particular cases of Fourier's series	648—656
444.	Dirichlet's integral	656—658
445—448.	Dirichlet's investigation of Fourier's series	658—666
449, 450.	Application of the second mean value theorem	666—669
451.	Uniform convergence of Fourier's series	669—670
452—455.	The limiting values of the coefficients in Fourier's series	671—678
456—457.	Sufficient conditions of convergence of Fourier's series at a point	678—683
458—461.	Conditions of uniform convergence of Fourier's series	683—695
32—464.	Further investigations of Dirichlet's integral	695—701
5—467.	The non-convergence of Fourier's series	701—707
—471.	The series of arithmetic means related to Fourier's series	707—714
472.	Properties of the coefficients of Fourier's series	715—717
73.	The integration of Fourier's series	718—719
74.	Properties of Poisson's integral	719—722
, 476.	Approximate representation of functions by finite trigonometrical series	722—725
77.	The differentiation of Fourier's series	725—728
	General Examples	728—730
—484.	Riemann's theory of trigonometrical series	730—742
485.	Investigations subsequent to those of Riemann	743—746
486.	The limits of the coefficients of a trigonometrical series	746—748
487.	Proof of the uniqueness of the trigonometrical series representing a function	748—749
8—490.	The representation of integrable functions	750—756
491.	The convergence of a trigonometrical series at a point	756—757
92, 493.	Fourier's integral representation of a function	758—762

APPENDIX.

On transfinite numbers and order-types . . . . . 763—766

LIST OF AUTHORS QUOTED . . . . . 767—768

GENERAL INDEX . . . . . 769—772

CORRIGENDA AND ADDENDA.

Page 96. In example 2, line 7, *for* "Of each rational number, there is a double representation" *read* "Of each rational number, not represented by a recurring radix-fraction, there is a double representation."

Page 267. In the statement of the theorem in § 203, *for* "*where θ is some proper fraction, and is neither* 0 *nor* 1" *read* "*where θ is such that* $0<\theta<1$." The number θ is not necessarily rational.

Page 268. In line 9, *for* "*for some value of θ which is a proper fraction, and is neither* 1 *nor* 0" *read* "*for some value of θ which is such that* $0<\theta<1$."

Page 317. The statement which commences on line 10 from the foot of the page is erroneous. The repeated limit may have a definite value when $\lim\limits_{k=0}\dfrac{f(x_0, y_0+k)-f(x_0, y_0)}{k}$ has no definite value. This is illustrated by the example $f(x, y)=\psi(x)+\chi(y)$, where $\psi(x)$, $\chi(y)$ are non-differentiable functions. In this case the above single limit does not exist, but the repeated limit exists, and is zero. The existence of the repeated limit as a definite number does not therefore necessarily imply the existence of $\dfrac{\partial f}{\partial y_0}$ and of $\dfrac{\partial^2 f}{\partial x_0 \partial y_0}$, the latter of these having been so defined that it only exists when $\dfrac{\partial f}{\partial y_0}$ exists. The point is more fully discussed in a paper by the Author in the *Proc. Lond. Math. Soc.*, ser. 2, vol. v, "On repeated limits."

Page 318. In the statement of the theorem, *for* "(3) $f(x, y)$ *be continuous with respect to* x *at* (x_0, y_0)," *read* "(3) $f(x, y)$ *be such that* $\dfrac{\partial f(x_0, y_0)}{\partial x_0}$, $\dfrac{\partial f(x_0, y_0)}{\partial y_0}$ *both exist.*"

Page 319. Line 5 from the foot of the page. Delete the sentence commencing "The existence of $\dfrac{\partial f}{\partial x_0}$, $\dfrac{\partial f}{\partial y_0}$."

Page 357. In the Example, line 8, *for* "let $F(x)=-\phi(x, \gamma)$" *read* "let $F(x)=-\phi(x, \beta)$."

Page 394. In line 8 from the foot of the page, *for* "in accordance with Riemann's definition" *read* "in accordance with Lebesgue's definition, the integrand having in each case only a finite number of values."

CHAPTER I.

NUMBER.

1. THE operation of counting, in which the integral numbers are employed, can be carried out by a mind to which discrete objects, which may be either physical or ideal*, are presented, and which possesses certain fundamental notions which we proceed to specify.

(1) The notion of *unity*, a form under which an object is conceived when it is regarded as a single one. An object so regarded may be either of a material or of a purely abstract or ideal nature, and may be recognized, for all other purposes than that of counting, as possessing any degree of complexity. It is sufficient, in order that the object may be regarded under the form of unity, that it be so far distinct from other objects, as to be recognized at the time when it is counted, as discrete and identifiable. What external marks are necessary that an object may be so recognized as discrete, is a matter for the judgment of the mind at the time when the object is counted. The unity under which the object is apprehended is a formal or logical, rather than a natural unity; it is more or less arbitrarily attributed to the object by the mind.

(2) The notion of a *collection* or *aggregate* of objects which is conceived of as containing more or fewer objects, or as possessing a greater or less degree of plurality. A group of objects regarded as an aggregate is conceived of not merely as a plurality of objects to each of which unity is ascribed as in (1), but also as itself an object to which unity is ascribed when it is regarded as a single whole. The single objects of which the aggregate is composed may be spoken of as the *elements* of the aggregate; such elements need not possess any parity as regards size or any other special quality, but may be of the most diverse characters: a certain logical parity is however

* It is held by some authors that the operation of counting is primarily applicable to physical objects only. Thus, J. S. Mill writes:—"The fact asserted in the definition of a number is a physical fact. Each of the numbers, two, three, four, etc., denotes physical phenomena, and connotes a physical property of these phenomena." See *Logic*, 9th edition, vol. II, p. 150. That objects which are not physical, can be counted, was maintained by Leibnitz and by Locke. See also Frege's *Grundlagen der Arithmetik*, Breslau, 1884, where an account is given of various views as to the nature and origin of the idea of Number.

ascribed to them in the process of counting, in virtue of the fact that each of them is regarded as a single object. A sensibly continuous presentation cannot be regarded as an aggregate containing a plurality of elements, until the mind has recognized in it sufficiently distinct lines of division to serve the purpose of marking off distinct objects within it, the totality of which makes up the whole presentation; for instance, the history of a country could be regarded as an aggregate of distinct periods, only when sufficiently salient features had been recognized in that history to warrant a judgment that periods were to be found in it, each of which had a sufficient degree of discreteness to be subsumed under the form of unity. In actual counting, the aggregate is not necessarily determinate before the counting is commenced, but becomes so when the process is completed; the notion of an aggregate is thus still necessary to the process of counting, if the process is ever to come to an end, or to be conceived of as having come to an end.

It has been held* that when an aggregate is counted, the elements must remain distinct from one another, not disappearing or combining with each other during the process. That this condition is unnecessary may be seen, for example, by considering the case of counting breakers on the sea-shore, or that of counting the vibrations of a pendulum; thus no physical permanence, but only an ideal one, is necessary.

A discussion of the characteristics which an aggregate (not necessarily finite) must possess, in order that it may be an object of mathematical thought, will be given in Chapter III.

(3) The notion of *order*, in virtue of which relative rank is given to each object in a collection, so that the collection becomes an ordered aggregate. In actual counting, the order is assigned to the objects during the process itself, as an order in time, and this may be done in an arbitrary manner; the order of the elements in an aggregate may, however, be assigned in a manner dependent upon their sizes, weights, or other qualities, or in accordance with their positions in space. Order may, however, be regarded as an abstract conception, independent of a particular mode of ordering; for an aggregate to be an ordered one, it is necessary that in some manner or other, each element be recognized as possessing a certain rank, in virtue of which it is known as regards any two elements which may be chosen, which of them has the lower, and which the higher rank. An element is said to precede any other element of higher rank than itself.

(4) The notion of *correspondence*, which underlies the process of tallying. The elements of one aggregate may be made to stand in some logical relation with those of another one, so that a definite element of one aggregate is regarded as correspondent to a definite element of another aggregate.

* See Helmholtz's *Zählen und Messen*, Leipzig, 1887; *Wissens. Abhandl.* vol. III, p. 372.

The correspondence may be complete, in the sense that to every element of either aggregate there corresponds one element, and one only, of the other aggregate; or the correspondence may be incomplete, in which case one of the aggregates has one or more elements to which no elements in the other aggregate correspond. In the latter case we say that the aggregate with the superfluous element or elements contains more elements than the other aggregate, and that the latter contains fewer elements than the former.

A correspondence between two aggregates is defined when specifications or rules are laid down which suffice to decide which element of one aggregate corresponds to each element of the other; so that, in the case of complete correspondence, no element of either aggregate is without a correspondent one in the other.

Whether, or how far, these fundamental notions of *unity, aggregate, order*, and *correspondence* should be regarded as derived empirically from experience, by a process of abstraction, or whether it must be held that they are original forms which the mind possesses prior to, and as the necessary conditions of the possibility of such experience, are questions into which it is beyond our province to enter. It is certain that civilized man possesses these fundamental notions, and it is highly probable that primitive man possessed them long before the notion of abstract number had appeared in an explicit and developed form. The investigation of the origin of these notions, and their further analysis is a matter for the Psychologist and for the Philosopher. Mathematical Science, as any other special science, must take its fundamental notions as data; it is concerned with the analysis of them, only so far as suffices to establish that they possess the degree of definiteness which such data must have, if they are to lie at the base of a logically ordered system.

ORDINAL NUMBERS.

2. If from an ordered aggregate some of the elements are removed, the aggregate which remains is said to be a *part* of the original aggregate. It will be observed that the relative order of any two elements in the part is the same as the relative order of those elements in the original aggregate.

An ordered aggregate is said to be *finite* when it satisfies the following conditions:—

(1) There is one element which has lower rank than any of the others.

(2) There is one element which has higher rank than any of the others.

(3) Every part of the aggregate has an element which has higher rank than every other element in the part, and also it has an element which has lower rank than any other element in the part.

These conditions are equivalent to the statement that a finite aggregate, and also each part of it, has a first and a last element.

Every part of a finite ordered aggregate is also a finite ordered aggregate.

If M be the aggregate, and M_1 a part of it, then M_1 has a highest and a lowest element; also every part of M_1 being also a part of M, has a lowest and a highest element; therefore M_1 is itself finite.

3. Two finite ordered aggregates are said to be *similar* when they can be made to completely correspond, so that to each element of either of them there corresponds a single element of the other, and so that to any two elements P, Q of the one there correspond two elements P', Q' of the other, which have the same relation as regards rank; viz. that if P is of lower rank than Q, then P' is of lower rank than Q', and if P is of higher rank than Q, then P' is of higher rank than Q'.

Two finite ordered aggregates which are similar are said to have the same *ordinal number.*

If each of two ordered aggregates is similar to a third, they are similar to one another. For if an element P of the first corresponds to an element R of the third, and the element Q of the second corresponds to R, it is clear that if we make P correspond to Q, the first two aggregates are made to correspond in such a way that the relative order is preserved.

It thus appears that *an ordinal number is characteristic of a class of similar ordered aggregates.*

An aggregate which consists of a single element A, is said to have the ordinal number one, denoted by the symbol 1. The ordinal number 1 is characteristic of every aggregate which consists of a single element.

If to the aggregate which consists of an element A, we adjoin a new element B, and assign to B a higher rank than A, we obtain an aggregate (A, B) which has an ordinal number 2, characteristic of all aggregates which are formed in this manner; A is said to be the first element, B the second. If to an ordered aggregate (A, B), of which the ordinal number is 2, we adjoin another element C, and regard this as having higher rank than A and B, we obtain an ordered aggregate (A, B, C), of which the ordinal number is called 3, and is characteristic of all ordered aggregates formed in this manner Proceeding in this way, if we have formed an ordered aggregate $(A, B, C \ldots\ldots H)$, of which the ordinal number is n, and adjoin to this aggregate a new element K, we obtain a new aggregate $(A, B, C, \ldots\ldots H, K)$, of which the ordinal number n' is different from n.

Every ordered aggregate which can be formed in the manner described is finite.

This can be proved by induction. Let us assume that M is a finite ordered aggregate: it will then be proved that (M, e), the ordered aggregate obtained by adjoining an element e of higher rank than the elements of M, is also finite. Since M has a lowest element, (M, e) has the same lowest element, also (M, e) has a highest element e. Again if M_1 is a part of (M, e) which does not contain e, then M_1 is a part of M, and therefore has a highest and a lowest element. If M_1 is a part of (M, e) which contains e, let it be (M_2, e), where M_2 is a part of M, and therefore contains a lowest element which is also the lowest element of (M_2, e); also (M_2, e) contains a highest element e. It has thus been shewn that (M, e) satisfies the requisite conditions that it should be finite, provided M does so. The aggregates $A, (A, B)$ are clearly finite: hence the method of induction proves that every ordered aggregate which can be formed by continually adjoining new elements to an aggregate which originally contained one element is a finite one.

Conversely, it can be shewn that *every finite ordered aggregate can be formed in the manner above described.*

Let M be a finite ordered aggregate, and let e' be its highest element, thus $M = (M_1, e')$. Now M_1 being a part of M, has a highest element e'', thus $M_1 = (M_2, e'')$, or $M = (M_2, e'', e')$. Proceeding in this manner, if we do not reach an aggregate M_r which contains a single element only, we shall have found a part $(\ldots\ldots e''', e'', e', e)$ of M which has no element of lowest rank. But this is impossible, since M is by hypothesis finite, and therefore contains no part without a lowest element. It has thus been shewn that M can be reduced, in the manner indicated, to an aggregate with a single element: and conversely, starting with this latter aggregate, M is obtained by adjoining to it successively new elements.

A finite ordered aggregate is not similar to any part of itself.

This theorem may also be proved by induction. For if we assume that the finite ordered aggregate M is not similar to any part of itself, it can be shewn that the same holds for (M, e). If possible let M_1 be a part of (M, e) which is similar to (M, e); then if M_1 contains e, it must be of the form (M_2, e), and if (M_2, e) is similar to (M, e), M_2 must be similar to M, which is contrary to the hypothesis that M contains no part similar to itself. If M_1 does not contain e, it must be of the form (M_2, f), where f is the element which corresponds to e in (M, e); in this case again M_2 is similar to M, and is a part of it; thus we have again a contradiction. The theorem holds for (A, B), and therefore generally.

It follows from this theorem that *the ordinal numbers* 1, 2, 3,...... which have been defined as the ordinal numbers of aggregates (A), (A, B), (A, B, C)...... *are all different from one another*, for each of these aggregates being a part of each of those which follow it, cannot be similar to any of the aggregates which follow it.

Each of the ordinal numbers is to be regarded as a unique ideal object

These conditions are equivalent to the statement that a finite aggregate, and also each part of it, has a first and a last element.

Every part of a finite ordered aggregate is also a finite ordered aggregate.

If M be the aggregate, and M_1 a part of it, then M_1 has a highest and a lowest element; also every part of M_1 being also a part of M, has a lowest and a highest element; therefore M_1 is itself finite.

3. Two finite ordered aggregates are said to be *similar* when they can be made to completely correspond, so that to each element of either of them there corresponds a single element of the other, and so that to any two elements P, Q of the one there correspond two elements P', Q' of the other, which have the same relation as regards rank; viz. that if P is of lower rank than Q, then P' is of lower rank than Q', and if P is of higher rank than Q, then P' is of higher rank than Q'.

Two finite ordered aggregates which are similar are said to have the same *ordinal number.*

If each of two ordered aggregates is similar to a third, they are similar to one another. For if an element P of the first corresponds to an element R of the third, and the element Q of the second corresponds to R, it is clear that if we make P correspond to Q, the first two aggregates are made to correspond in such a way that the relative order is preserved.

It thus appears that *an ordinal number is characteristic of a class of similar ordered aggregates.*

An aggregate which consists of a single element A, is said to have the ordinal number one, denoted by the symbol 1. The ordinal number 1 is characteristic of every aggregate which consists of a single element.

If to the aggregate which consists of an element A, we adjoin a new element B, and assign to B a higher rank than A, we obtain an aggregate (A, B) which has an ordinal number 2, characteristic of all aggregates which are formed in this manner; A is said to be the first element, B the second. If to an ordered aggregate (A, B), of which the ordinal number is 2, we adjoin another element C, and regard this as having higher rank than A and B, we obtain an ordered aggregate (A, B, C), of which the ordinal number is called 3, and is characteristic of all ordered aggregates formed in this manner Proceeding in this way, if we have formed an ordered aggregate $(A, B, C \ldots\ldots H)$, of which the ordinal number is n, and adjoin to this aggregate a new element K, we obtain a new aggregate $(A, B, C, \ldots\ldots H, K)$, of which the ordinal number n' is different from n.

Every ordered aggregate which can be formed in the manner described is finite.

This can be proved by induction. Let us assume that M is a finite ordered aggregate: it will then be proved that (M, e), the ordered aggregate obtained by adjoining an element e of higher rank than the elements of M, is also finite. Since M has a lowest element, (M, e) has the same lowest element, also (M, e) has a highest element e. Again if M_1 is a part of (M, e) which does not contain e, then M_1 is a part of M, and therefore has a highest and a lowest element. If M_1 is a part of (M, e) which contains e, let it be (M_2, e), where M_2 is a part of M, and therefore contains a lowest element which is also the lowest element of (M_2, e); also (M_2, e) contains a highest element e. It has thus been shewn that (M, e) satisfies the requisite conditions that it should be finite, provided M does so. The aggregates $A, (A, B)$ are clearly finite: hence the method of induction proves that every ordered aggregate which can be formed by continually adjoining new elements to an aggregate which originally contained one element is a finite one.

Conversely, it can be shewn that *every finite ordered aggregate can be formed in the manner above described.*

Let M be a finite ordered aggregate, and let e' be its highest element, thus $M = (M_1, e')$. Now M_1 being a part of M, has a highest element e'', thus $M_1 = (M_2, e'')$, or $M = (M_2, e'', e')$. Proceeding in this manner, if we do not reach an aggregate M_r which contains a single element only, we shall have found a part $(\ldots\ldots e''', e'', e', e)$ of M which has no element of lowest rank. But this is impossible, since M is by hypothesis finite, and therefore contains no part without a lowest element. It has thus been shewn that M can be reduced, in the manner indicated, to an aggregate with a single element: and conversely, starting with this latter aggregate, M is obtained by adjoining to it successively new elements.

A finite ordered aggregate is not similar to any part of itself.

This theorem may also be proved by induction. For if we assume that the finite ordered aggregate M is not similar to any part of itself, it can be shewn that the same holds for (M, e). If possible let M_1 be a part of (M, e) which is similar to (M, e); then if M_1 contains e, it must be of the form (M_2, e), and if (M_2, e) is similar to (M, e), M_2 must be similar to M, which is contrary to the hypothesis that M contains no part similar to itself. If M_1 does not contain e, it must be of the form (M_2, f), where f is the element which corresponds to e in (M, e); in this case again M_2 is similar to M, and is a part of it; thus we have again a contradiction. The theorem holds for (A, B), and therefore generally.

It follows from this theorem that *the ordinal numbers* 1, 2, 3,...... which have been defined as the ordinal numbers of aggregates (A), (A, B), (A, B, C)...... *are all different from one another*, for each of these aggregates being a part of each of those which follow it, cannot be similar to any of the aggregates which follow it.

Each of the ordinal numbers is to be regarded as a unique ideal object in that it is a permanent object for thought. The relation of an ordinal number to an ordered aggregate of objects which is characterised by that number, may be illustrated by the analogy of the relation between *the colour* red, and a particular red object.

4. *A simply infinite ascending aggregate,* or *simple sequence,* is an ordered aggregate which has no element of higher rank than all the others, and is such that every part which has an element of higher rank than all the other elements in that part, is a finite ordered aggregate.

It follows from this definition, that in a simple sequence there is one element of lower rank than all the others; and further, that every part of the simple sequence has an element of lower rank than all the other elements in that part.

A simply infinite ascending aggregate differs from a finite ordered aggregate in having no element which is of higher rank than all the other elements.

The totality of ordinal numbers forms a simply infinite ascending aggregate; these objects may be represented by a set of signs

$$\alpha, \beta, \gamma, \delta, \ldots$$

or

$$1, 2, 3, 4, \ldots$$

where it is assumed that some adequate scheme of such signs has been devised.

The order of the elements is assigned by the successive formation, as above, of aggregates having the various elements for their ordinal numbers, and it has been shewn that if an aggregate has the ordinal number n, another aggregate having a different ordinal number n', taken to be of next higher rank than n, can be formed. There exists therefore no highest ordinal number.

Instead of using the expressions "of higher rank" and "of lower rank," it is usual to say that a number m is less than a number n, when m is of lower rank than n in the ordered aggregate of ordinal numbers, and that n is greater than m. The terms "greater" and "less" are borrowed from the language primarily applicable to the description of magnitudes: but in pure arithmetic and pure analysis generally, they are used only in the sense in which they indicate higher or lower rank, and this rank has no necessary reference to relations of magnitude or of measurable quantity.

The operation of counting a finite aggregate of objects of any kind may be conceived of as the process of putting the objects into correspondence with the elements of the aggregate of ordinal numbers, in such a way, that when any ordinal number has an element of the aggregate which corresponds to it, each of the preceding ordinal numbers also has an element which corresponds to it. The finite aggregate is usually ordered by the process itself, the ranks of the various elements being successively assigned to them as the counting proceeds. Those ordinal numbers which are employed in counting such an aggregate may be regarded as forming an aggregate which is similar to the given aggregate, as ordered by the process of counting. The last of the ordinal numbers employed in counting a finite aggregate, is *the ordinal number*, or simply *the number* (Anzahl) of the ordered aggregate.

The theorem that an ordered aggregate is not similar to any of its parts, holds only as regards finite aggregates. It will appear in the course of the discussion in Chapter III. that every aggregate which is not finite has parts which are similar to the whole; and this property is sometimes taken as the basis of the definition of an infinite, or transfinite, aggregate. For example, the aggregate of ordinal numbers 1, 2, 3, ... is similar to the part 2, 4, 6, ... which contains the even numbers only.

CARDINAL NUMBERS.

5. *If any finite ordered aggregate be re-ordered in any manner, the new ordered aggregate is finite, and has the same ordinal number as the original one.*

In order to prove this theorem, the following particular case will be first established:—If Q is a finite ordered aggregate, the aggregate (Q, e) obtained by adjoining to Q a new element e of higher rank than all the elements of Q, is similar to (e, Q), in which e has a lower rank than all the elements of Q. For let $Q \equiv (Q_1, f)$, and let us assume that the theorem holds for Q_1, *i.e.* that (Q_1, e) is similar to (e, Q_1); it follows, since a complete correspondence can be established between the elements of (Q_1, e) and (e, Q_1), that the same is true of the two aggregates (Q_1, e, f) and (e, Q_1, f). Now (Q_1, e, f) is similar to (Q_1, f, e), since Q_1 can be made to correspond to itself, e to f, and f to e, therefore (Q_1, f, e) is similar to (e, Q_1, f), or (Q, e) to (e, Q), and thus the theorem holds for $Q \equiv (Q_1, f)$, provided it holds for Q_1. Now it clearly holds if Q_1 consists of a single element; hence by induction it holds for any finite ordered aggregate Q. To prove the theorem in the general case, let us assume that it is true for an aggregate M; it will then be shewn to be true for (M, e). For let an aggregate obtained by re-ordering (M, e) be (R, e, S), where either R or S may be absent; (R, e, S) is similar to (R, S, e), for R corresponds with itself, and it has been shewn above that (e, S) is similar to (S, e). Since (R, S) is by hypothesis similar to M, it follows that (R, S, e) is similar to (M, e), and therefore (R, e, S) is similar to (M, e). The theorem clearly holds for an aggregate (A, B) which contains two elements, hence by induction it holds for every finite ordered aggregate.

It follows from the theorem which has been established above, that, *for any aggregate which can be ordered as a finite ordered aggregate, the ordinal number is independent of the mode in which the aggregate is ordered.*

It will be found, when the generalization of ordinal numbers for non-finite aggregates is considered in Chapter III., that this property, that the ordinal number of an aggregate is independent of the mode of ordering, is peculiar to finite aggregates.

6. Two aggregates are said to be *equivalent*, when their elements can be placed into correspondence so that to each element of either aggregate there corresponds one and only one element of the other aggregate.

It will be observed that the relation of equivalence differs from that of similarity in that it contains no reference to order. It is clear that two aggregates which are each equivalent to a third, are equivalent to one another.

An unordered aggregate is said to be *finite*, when it can be so ordered that the ordered aggregate is finite in accordance with the definition given in § 2.

Two (finite) aggregates which are equivalent are said to have the same *cardinal number*.

It thus appears that *a cardinal number is characteristic of a class of equivalent aggregates.*

Each of the cardinal numbers is to be regarded as a unique ideal object; the relation of a cardinal number to a member of the class of equivalent aggregates of objects, of which it is characteristic, may be illustrated in the same manner as in § 3, in the case of the ordinal numbers.

Since all similar aggregates are also equivalent, and since, in the case of a finite aggregate, the ordinal number is independent of the mode in which the aggregate is ordered, it follows that for every finite ordinal number there is a corresponding cardinal number.

The cardinal numbers of finite aggregates are denoted by the same symbols 1, 2, 3, ... as the corresponding ordinal numbers. The two kinds of numbers are not symbolically distinguished from each other, although logically they are not identical.

It will be seen in Chapter III. that this practical identity of ordinal and cardinal numbers is confined to the case of the numbers corresponding to finite aggregates, and therefore called finite numbers. The finite cardinal numbers form a simple sequence 1, 2, 3, ... similar to the sequence of finite ordinal numbers; the expressions "greater" and "less" are used in relation to two cardinal numbers in the same purely ordinal sense, denoting higher and lower rank, as in the case of ordinal numbers.

It is impossible, in a purely mathematical work, to enter into a discussion of the nature and proper definition of number from a philosophical point of view. One view of number which is widely held, is embodied in the definition by abstraction, in which the cardinal number* is regarded as the concept of an aggregate which remains when we make abstraction of the nature of the objects forming the aggregate, and of the order in which they are given; the ordinal number is then regarded as the concept obtained by making abstraction of the nature of the objects only, retaining the order† in which they are given in the aggregate. The view has also been maintained‡ that a cardinal number is simply the class of all equivalent aggregates. A tendency has been exhibited amongst mathematicians§ to regard numbers, at least for the purposes of analysis, as identical with the symbols which represent them. In accordance with this view, abstract arithmetic is cut entirely adrift from the fundamental notions related to experience in which it had its origin, and it is thus reduced to a species of mechanical game played in accordance with a set of rules which, when divorced from their origin, have the appearance of being perfectly arbitrary; though it may, of course, be said that it is possible at the end of any arithmetical process to reconnect the symbols employed, with the ideas which originally suggested them, and thus to interpret the results of the purely symbolical processes. Whatever view‖ be adopted as to the real nature of number and its place in a general

* This view is that of G. Cantor; see *Math. Annalen*, vol. XLVI, p. 481, where the following definition is given:—"'Mächtigkeit,' oder 'Cardinalzahl' von *M* nennen wir den Allgemeinbegriff welcher mit Hülfe unseres activen Denkvermögens aus der Menge *M* hervorgeht, dass von der Beschaffenheit ihrer verschiedenen Elemente *m*, und von der Ordnung ihres Gegebenseins abstrahirt wird." See also Peano, *Formulaires de Mathématiques*, 1901, § 32, ·0 Note.

† Ordinal numbers are frequently regarded as logically prior to cardinal numbers, but this order of procedure is not a necessary one. In Dedekind's tract "Was sind und was sollen die Zahlen," Brunswick, 1887 and 1893, which has been translated into English by Prof. W. W. Beman, under the title "Essays on the Theory of Numbers," 1901, a detailed treatment of the subject is given, in which the notion of order is regarded as fundamental.

‡ See B. Russell, *The Principles of Mathematics*, vol. I, chap. XI.

§ For example see Heine, *Crelle's Journal*, vol. LXXIV (1872), p. 173, where the matter is stated in the following plain form: "Ich nenne gewisse greifbare Zeichen Zahlen, sodass die Existenz dieser Zahlen also nicht in Frage steht." Again, Helmholtz appears to hold a view closely approaching the notion that Arithmetic is the art of manipulating certain signs according to certain rules of operation; he writes in *Ges. Abh.* vol. III, p. 359, "Ich betrachte die Arithmetik oder die Lehre von den reinen Zahlen als eine auf rein psychologische Thatsachen aufgebaute Methode, durch die die folgerichtige Anwendung eines Zeichensystems (nämlich der Zahlen) von unbegrenzter Ausdehnung und unbegrenzter Möglichkeit der Verfeinerung gelehrt wird." Reference may be made to an essay by A. Pringsheim in the *Jahresberichte der d. math. Vereinigung*, vol. VI, 1899, "Ueber den Zahl- und Grenzbegriff im Unterricht." In an article entitled "Die Du Bois Reymond'sche Convergenz-Grenze," *Sitzungsberichte d. bayer. Akad.* vol. XXVII, 1897, Pringsheim speaks of numbers as "Zeichen, denen lediglich eine bestimmte Succession zukommt." See p. 326. This article contains various remarks on arithmetization, and especially a criticism of the views of P. Du Bois Reymond. A searching criticism of the tendency to reduce Arithmetic to the formal manipulation of symbols is given in L. Couturat's work *De l'infini mathématique*, Paris, 1896, which contains a valuable account and discussion of theories of the philosophy of arithmetic.

‖ References to the literature relating to the Philosophy of Number will be found in the Article I. A. 1, "Grundlagen der Arithmetik," by H. Schubert, in the *Encyclopädie der mathe-*

scheme of thought, the assumption of the right to hypostatize numbers would appear to be an essential condition of the possibility of developing an abstract arithmetic, and consequently of the establishment of mathematical analysis in general.

THE OPERATIONS ON INTEGRAL NUMBERS.

7. If two finite ordered aggregates A and B, of which the ordinal numbers are a and b respectively, are combined into a single ordered aggregate in which the elements of A have all lower rank than those of B, and in which any two elements of A, and any two elements of B, have the same relative orders as in the original aggregates, *then the ordinal number of the combined aggregate is said to be the sum of the ordinal numbers a and b*, and is denoted by $a+b$.

It can be shewn that the new aggregate is a finite one, and that its ordinal number is unaltered if for A and B there be substituted aggregates which are similar to them; it thus appears that the sum $a+b$ is a finite number which depends only upon a and b.

The aggregate (A, B) has as lowest element the lowest element of A, and as highest element the highest element of B; moreover any part of (A, B) is of the form (A', B'), where A' is a part of A, and B' is a part of B; or else it has one of the forms A', B', and since A', B' have each a lowest and a highest element, any such part of (A, B) has a lowest and a highest element. Thus (A, B) is finite.

Again, if A_1, B_1 are aggregates which are similar to A and B respectively, the elements of A may be placed in correspondence with those of A_1, and the elements of B with those of B_1; we have then a $(1, 1)$ correspondence between the elements of (A, B) and those of (A_1, B_1); thus the ordinal number of (A, B) is the same as that of (A_1, B_1).

Since (A, B) has the same ordinal number as (B, A) it follows that $a+b=b+a$, which is known as the commutative law of addition.

If a, b are the cardinal numbers of two finite aggregates A, B, then the cardinal number of the aggregate formed by combining the two aggregates into one is said to be the *sum* of a and b, and is denoted by $a+b$. That $a+b$ is a definite finite number dependent only on a and b, follows at once from the corresponding theorem which has been proved for cardinal numbers.

matischen Wissenschaften, vol. I; also in E. G. Husserl's *Philosophie der Arithmetik*, vol. I, chaps. 5 and 6, Halle, 1891. The view that Number is fundamentally dependent on the notion of Time was developed by Sir W. R. Hamilton; see the *Dublin Transactions*, vol. XVII (II), 1835, "Theory of Conjugate Functions or Algebraic Couples with a Preliminary and Elementary Essay on Algebra as the Science of Pure Time"; see also Helmholtz's essay "Zählen und Messen" (1887), where the view is adopted that the axioms of Arithmetic have a relation to the intuitional form of Time, similar to that which the axioms of Geometry have to the intuitional form of space.

The operation of finding the sum of two numbers a and b, is known as the operation of addition, and it has been shewn that this operation is commutative. It should be observed that the sum of two numbers a and b cannot be determined merely by contemplating those numbers themselves as abstract concepts, but can only be defined as above, by referring to aggregates of which a and b are the numbers, and then combining those aggregates. The number of the combined aggregate is then conceived of as the result of a symbolical operation upon the numbers a and b. For example, the equation $5+3=8$ does not imply that the concept 8 is obtainable by placing the concepts 5, 3 as it were in juxtaposition, but can only be regarded as a symbolical expression of the fact that an aggregate of 5 objects together with one of 3 objects make up an aggregate of 8 objects. Bearing this observation in mind, the numbers 1, 2, 3,... are represented symbolically as the results of successive operations of addition, $1+1=2$, $2+1=3$, $3+1=4$, etc.; but these equations do not express definitions of the numbers 2, 3, 4,..., since from the concept unity taken by itself, no other concept is directly derivable.

The operation of addition can be extended by continued repetition. Thus the sum of $a, b, c, \ldots k$ is a finite number represented by $a+b+c+\ldots+k$, and, in particular, any number n is represented by $n=1+1+1+\ldots+1$. An immediate induction shews that the result of the operation of addition repeated any definite number of times is a finite number dependent only on the constituents of the summation.

The associative law of addition, $a+(b+c)=(a+b)+c$, follows from the irrelevancy of the order in which the operations are performed. This is seen from the contemplation of aggregates of which a, b, c are either the ordinal or the cardinal numbers.

8. If in a finite aggregate of which the number is b, each element be replaced by a finite aggregate of which the number is a, *the number of the new aggregate so formed is said to be the product of b by a*, and is denoted by ab. This operation is said to be that of multiplying b by a. By taking the aggregates to be ordered, it is seen at once that the new aggregate satisfies the conditions that it is finite, and that its number is unaltered by the substitution of similar aggregates of other objects for those originally employed. Thus ab is a definite number dependent only on a and b.

It is clear that ab may be regarded as the sum $a+a+a+\ldots$, where a occurs b times in the operation.

If the ordered aggregate of which the number is ab, be re-ordered in the following manner:—take the first element of each of the aggregates of which a is the number, then the second elements of these aggregates, and so on, with lastly the ath elements of these aggregates, then we have as the result of the process an aggregate of which the number is a, and each element of which consists of an aggregate of which the number is b; the re-ordered aggregate has the number ba. It has thus been shewn that $ab = ba$, which is expressed by saying that the operation of multiplication of finite integers is commutative.

The distributive law for multiplication, $a(b+c) = ab + ac$, follows from the definition of the operation, by considering the aggregates of which a, b, c are the numbers.

An immediate induction shews that the repetition of the operation of multiplication any definite number of times gives a finite number dependent only on the numbers multiplied, and independent of the order in which the operations are performed.

The result of the operation of multiplying the number a by itself is denoted by a^n, where n is the number of times a occurs in the product $a.a.a \dots a$. From this definition the law $a^m . a^n = a^{m+n}$, is directly deducible.

9. If the sum of two numbers a, b be denoted by c, the number a is uniquely determined when b, c are fixed; and it is then regarded as the result of the operation of subtracting b from c. The operation of subtraction is thus defined as inverse to that of addition. If $c = a + b$, a is obtained as the result of the operation denoted by $c - b$, which is such that $(c-b) + b = c$. It is obvious that the operation of subtraction of b from c is only possible in case $c > b$.

If the product of two numbers a, b be the number c, then the number a is uniquely determined when b and c are given; and a is regarded as the result of the operation of division of c by b. The operation of division so defined is inverse to that of multiplication; it is clear that the operation is only possible in case c is one of the class of numbers b, $2b$, $3b$, ...

FRACTIONAL NUMBERS.

10. The operation of multiplying two integers a, b together, is one which is always a possible operation, in accordance with the definition of the operation of multiplication which has been given above; the inverse operation of division is however, as we have seen, not always a possible one. This restriction upon the possibility of the operation of division suggests the introduction into Arithmetic of a new class of numbers, the rational fractions, which, when defined, shall be such that the operation of division, within the whole aggregate of integers and fractions, may be a possible one without restriction. Stated in algebraical form, the demand arises for a scheme of numbers such that the equation $ax = b$, shall always have a solution in x, where a, b are any two numbers which belong to the contemplated aggregate of numbers.

The actual use of fractional numbers arose historically from the necessities of the process of measurement of extensive magnitude, and the conception of a fraction which arises in this connection is the one which is used in ordinary life, and is made the basis of the treatment of the theory of fractions, even in recent scientific text-books. In accordance with this view, a unit of magnitude of some kind is divided into b equal parts, and a of these parts are taken; the resulting magnitude is then denoted by the fraction a/b.

This notion of the essential nature of a fraction, dependent as it is upon the notions of a *unit*, and of the divisibility of such unit into *equal parts*, is incompatible with the modern view that Mathematical Analysis should be developed upon the basis of a Pure Arithmetic, quite independently of all notions connected with the measurement of extensive magnitude. The modern tendency known as Arithmetization manifests itself in the construction of theories of Number and of the operations involving numbers, which depend entirely upon the conceptions connected with the process of counting, measurement being regarded as a process foreign to Pure Arithmetic. The process of counting is an exact one: whereas measurement can in practice only be carried out with a greater or less degree of approximation, and can only ideally be made an exact process. Pure Arithmetic is made the basis of Analysis, not only in accordance with the general principle that the fundamental conceptions of a branch of science should be irreducible to simpler conceptions, but also because the theory of ideally exact measurement has peculiar difficulties of its own. Our essentially inexact intuitions of spatial, temporal, or other magnitudes, necessitate a process of idealization in which the objects of perception are replaced by ideal objects subject to an exact scheme of definitions and postulates, in order that an exact science of measurement may be possible. The view is at present held by the majority of mathematicians that the nature of the abstract continuum, and that of a limit, are capable of exact formulation only in the language of a Pure Arithmetic; and that this science must therefore be developed upon an independent basis before it can be applied to the elucidation of the conceptions requisite for an abstract theory of continuous magnitude. The theory of measurement is, in accordance with this view, regarded as an application, and not as part of the basis, of Mathematical Analysis.

11. By those writers who are under the influence of the modern arithmetizing tendency, the traditional non-arithmetical definition of a fraction has been abandoned, and in its place a formal definition has been substituted, in which the fraction is regarded as an association of a pair of integers. The associated integers are regarded as making a single object, and laws of combination of these objects are then postulated.

If a, b are two integers, a new number (a, b), or in ordinary notation

$\frac{a}{b}$, is formed by the association of a and b, the new number being defined to be such as to satisfy the following conditions:

(1) (a, b) is regarded as ordinally greater, equal to, or less than (c, d), according as ad is greater, equal to, or less than bc. The expressions greater, equal to, or less than, are here used, not in their primitive sense as referring to magnitude, but in the sense in which we have used them in the case of integers, as assigning relative order to the numbers.

(2) $(a, 1)$ is defined as equal to a; thus if $b=1$, the association is regarded as equivalent to the integer a. Taking (1) in conjunction with this postulate, the new numbers have their orders assigned, not only relatively to one another, but relatively also to the integral numbers; so that the whole aggregate of integers and fractions is ordered, in the sense that, of two given numbers, it can always be said which has the higher rank.

(3) The addition of two fractional numbers is defined by

$$(a, b) + (c, d) = (ad + bc, bd).$$

(4) The multiplication of fractional numbers is defined by

$$(a, b) \times (c, d) = (ac, bd).$$

(5) The use of a fraction as an index, is defined by the postulate

$$x^{(a, b)} \times x^{(c, d)} = x^{(a, b)+(c, d)},$$

where x is any number, either integral or fractional. The symbol $x^{(a, b)}$ is to be interpreted subject to this postulate, in case such interpretation is possible.

It will be observed that, in the case $b=1$, $d=1$, the above definitions are consistent with those which have been adopted in the case of integral numbers; and thus the new numbers, together with the integers, form an aggregate with uniform laws of operations. It is easily seen that the operations with new numbers satisfy the commutative, associative, and distributive laws. The inverse operation of division is now one which is always possible within the domain of the numbers; thus $(a, b) \div (c, d) = (ad, bc)$. The inverse operation of subtraction, $(a, b) - (c, d) = (ad - bc, bd)$, is only possible if $(a, b) > (c, d)$.

The association of a pair of integers is a "number" in quite a different sense from that in which the cardinal and ordinal numbers, hitherto discussed, are numbers. The justification of the extension of the term "number" to the fractions, lies in the fact that a consistent scheme of operations can be imposed upon them, of which the laws are in agreement with those which hold for operations which involve integers only.

12. The scheme which has been above indicated suffices for a formal definition and logical development of the properties of fractions, but it is subject to the objection that it is of an arbitrary character; indeed it is not easy to see why the particular laws of operations have been postulated, except as suggested by the traditional non-arithmetical conception of a fraction.

To remedy this defect, a view of the nature of a fraction will be here given which relates the fraction with the process of counting, in such a manner that fractional and integral numbers have similar relations to that process. It will appear that the laws of combination given above naturally follow from this mode of regarding the fraction, with the exception of (5), which is however immediately suggested by the rule for integral indices.

Consider an aggregate of b objects, and out of these b objects pick out any a ($\leqq b$) of them. If we regard these a objects not only as single objects of number a, but also as belonging to an aggregate whose number is b, we may denote the a objects by (a, b), where their number a is associated with the cardinal number b of the aggregate to which they belong. This process being independent of the particular aggregate used, the abstract fraction (a, b) is related to this process in an analogous manner to that in which the number b is related to the process of counting an aggregate whose cardinal number is b. Thus the fraction (a, b), or a/b, is characteristic of an aggregate of a objects each of which belongs to an aggregate of b objects. The extension of the definition to the case $a > b$, is clear when we observe that it is unessential that the a objects taken should all belong to one and the same aggregate of b objects; it is sufficient that each of them be regarded as essentially belonging to *some* aggregate of cardinal number b. In accordance with this view, a fraction, say $3/5$, is characteristic of any three things each of which belongs to an aggregate of five things, *i.e.* $3/5$ means 3 out of 5. That the three things taken out of five should necessarily all be equal in respect of size, or some other kind of magnitude, is as irrelevant to the true nature of a fraction as the assumption of five things necessarily meaning five equal things, is to the true nature of the number five.

Since $(a, 1)$ is characteristic of an aggregate of a things each of which is also regarded as a single object, it is clear that $(a, 1)$ is identical with a.

If we suppose each of the b elements in an aggregate, of which the cardinal number is b, to be replaced by an aggregate of n elements, we have now an aggregate with nb for its cardinal number; and instead of a elements chosen out of this aggregate we now have na of the new elements, each of which is to be regarded as associated with the cardinal number nb. We represent these na elements by (na, nb), which is equivalent to (a, b), since the two forms represent two different aspects of the same process.

Therefore we have $(a, b) = (na, nb)$, or in the ordinary notation $a/b = na/nb$. This relation is in complete accordance with the law of logical (not arithmetical) addition, that a mere repetition of a term yields only the term itself.

Since $(a, b) = (ad, bd)$, and $(c, d) = (bc, bd)$, we regard (a, b) as greater, equal to, or less than (c, d), in the purely ordinal sense of the terms, according as ad is $\gtreqless bc$. For the two numbers (ad, bd), (bc, bd) are characteristic of the process of taking ad, bc elements respectively from an aggregate of the same cardinal number bd; and thus the relative order of the two numbers (a, b), (c, d) will naturally be fixed in accordance with the relative order of the two numbers ad, bc.

The addition of the two numbers (a, b) and (c, d) is equivalent to that of (ad, bd) and (bc, bd), and is consequently naturally defined as given by $(ad + bc, bd)$, which characterises the amalgamation of two aggregates of which the numbers are ad, bc, the elements of each of which all belong to an aggregate of number bd, or to one of several such aggregates.

To interpret the operation of multiplication, let us consider an object represented by (c, d); this consists of c things each belonging to an aggregate of d things. To multiply it by (a, b), is to take a such objects each of which belongs to an aggregate of b such objects; we have on the whole one or more aggregates of bd elements, and out of these, ac elements are to be taken. Thus the multiplication of the number (c, d) by the number (a, b) may be understood to characterise the result of taking a objects each of which is characterised by (c, d), out of one or more collections of b objects each of which objects is characterised by (c, d). This is the same thing as the process of taking ac objects out of one or more aggregates of bd objects, and is characterised by the number (ac, bd); we are thus led to the law of multiplication

$$(a, b) \times (c, d) = (ac, bd), \text{ or } \frac{a}{b} \times \frac{c}{d} = \frac{ac}{bd}.$$

NEGATIVE NUMBERS, AND THE NUMBER ZERO.

13. Although the operation of addition is always possible within the aggregate of integral and fractional numbers, yet the inverse operation of subtraction is not always possible; thus a number x cannot be found such that $x + (c, d) = (a, b)$, unless $(a, b) > (c, d)$. As the limitation of the possibility of division suggests the introduction of fractional numbers, so this limitation of the possibility of subtraction suggests the introduction of a further set of new numbers, which shall be such that within the so completed aggregate, subtraction may always be a possible operation.

If $\alpha, \beta, \gamma, \delta$ denote integral or fractional numbers such that $\alpha > \beta, \gamma > \delta$; we may put $\alpha = \beta + x, \gamma = \delta + y$; then $x = \alpha - \beta, y = \gamma - \delta$. We have

$$\alpha + \gamma = \beta + \delta + x + y,$$

hence $$x + y = (\alpha + \gamma) - (\beta + \delta),$$

or $$(\alpha - \beta) + (\gamma - \delta) = (\alpha + \gamma) - (\beta + \delta) \quad \text{.................(1).}$$

Again, if $\alpha - \beta = \gamma - \delta$, *i.e.* $x = y$, we have $\alpha + \delta = \beta + \delta + x = \beta + \gamma$;

or $$\alpha + \delta = \beta + \gamma, \text{ if } \alpha - \beta = \gamma - \delta \quad \text{....................(2).}$$

Lastly, we have

$$\alpha\gamma = (\beta + x)(\delta + y) = \beta(\delta + y) + x(\delta + y)$$
$$= \beta\delta + \beta y + x\delta + xy;$$

hence $$\alpha\gamma + \beta\delta = \beta(y + \delta) + \delta(x + \beta) + xy$$
$$= \beta\gamma + \alpha\delta + xy;$$

hence $$(\alpha - \beta)(\gamma - \delta) = (\alpha\gamma + \beta\delta) - (\alpha\delta + \beta\gamma) \quad \text{.................(3).}$$

The rules (1), (2), (3), with regard to the numbers $\alpha - \beta, \gamma - \delta$, which so far exist only when $\alpha > \beta, \gamma > \delta$, suggest the mode of the extension referred to above.

14. Let α, β be any two numbers integral or fractional, and conceive a new number $D(\alpha, \beta)$, formed by the association of α and β, to be defined as subject to the laws

(4) $D(\alpha, \beta) = D(\gamma, \delta)$, if $\alpha + \delta = \beta + \gamma$,

(5) $D(\alpha, \beta) + D(\gamma, \delta) = D(\alpha + \gamma, \beta + \delta)$,

(6) $D(\alpha, \beta) \times D(\gamma, \delta) = D(\alpha\gamma + \beta\delta, \alpha\delta + \beta\gamma)$:

it will be observed that when $\alpha > \beta$, and $\gamma > \delta$, $D(\alpha, \beta)$ may denote $\alpha - \beta$, the three laws becoming (2), (1), (3). It will now be shewn that the symbol $D(\alpha, \beta)$ defines a number of an aggregate within which the operation of subtraction is always possible. For, to find a number x, such that

$$x + D(\alpha, \beta) = D(\gamma, \delta), \text{ we see that } x = D(\beta + \gamma, \alpha + \delta),$$

since $$D(\alpha + \beta + \gamma, \alpha + \beta + \delta) = D(\gamma, \delta), \text{ in virtue of (4).}$$

Since $D(\alpha, \alpha) = D(\gamma, \gamma)$, we see that $D(\alpha, \alpha)$ is independent of α; and thus $D(\alpha, \alpha)$ defines a new number which is called the *number zero, and is denoted by the symbol* 0.

The number zero is regarded as characteristic of the absence of all elements from an aggregate of which the existence has been contemplated; it is the number of such a hypothetical aggregate, in a sense similar to that in which a positive integer is the number of an actual aggregate.

The number $D(\alpha + k, k)$ depends only on α, and we shall postulate that it is identical in meaning with α itself.

The numbers $D(\alpha, \beta)$, or $\alpha - \beta$, for which $\alpha > \beta$, are called positive numbers, and form the aggregate of integral and fractional numbers we have previously considered.

Those numbers for which $\alpha < \beta$, are called negative numbers.

Since by (5), $D(\alpha, \beta) + D(\beta, \alpha) = D(\alpha + \beta, \alpha + \beta) = 0$, the number $D(\beta, \alpha)$ may be denoted by $-D(\alpha, \beta)$, or in ordinary notation $-(\alpha - \beta)$. Thus *to every positive number x there corresponds a single negative number $-x$, which is such that $x + (-x) = 0$.*

We may now use the notation $\alpha - \beta$ in every case for $D(\alpha, \beta)$, and thus

$$\alpha - \beta = -(\beta - \alpha).$$

From (6), it is seen that the operation of division is always possible for two members of the complete aggregate of positive and negative numbers, and zero, except when the divisor is the number zero, in which case the operation is meaningless.

From (6), we see by putting $\gamma = \delta$, $N = D(\alpha, \beta)$, that $N.0 = 0$. From (5), we have $N + 0 = N$.

Any number $D(\alpha, \beta)$ is said to be greater in the ordinal sense than $D(\gamma, \delta)$, when $D(\alpha, \beta) - D(\gamma, \delta)$ is positive; thus the complete aggregate of positive and negative integral and fractional numbers together with the number zero, is one in which all the numbers are arranged in a definite order. This aggregate is known as the *aggregate of rational numbers.*

In the aggregate of rational numbers so ordered, the number zero has lower rank than any of the positive numbers, and higher rank than any of the negative numbers. Further, if x, y are two positive numbers of which x has higher rank than y, the negative number $-x$ has lower rank than $-y$.

If x, y are any two rational numbers, such that $x < y$, there exist an unlimited number of rational numbers each of which is $> x$, and $< y$.

Such numbers are said to be *between* x and y. For it can be seen at once, from the definition of order given above, that $\frac{1}{2}(x + y)$ is one such number; between $\frac{1}{2}(x + y)$ and either x or y, another rational number can, in a similar manner, be found. This process can be carried on without end; and it is clear that in accordance with the mode of ordering of the aggregate, defined above, all the numbers thus determined are between x and y.

If x, y are any two positive rational numbers such that $x < y$, an integer n can be found which is such that $nx > y$.

For $\frac{y}{x}$ is a rational number such that $\frac{y}{x}.x = x$; again if $\frac{p}{q}$ be any positive rational number, there exist integers which are $> \frac{p}{q}$, for $p + 1$ is itself such an integer. If n is an integer which is $> \frac{y}{x}$, we have $nx > \frac{y}{x}.x > y$; thus the theorem is established.

IRRATIONAL NUMBERS.

15. The only numbers of which the existence was recognized by the Greek geometers were the rational numbers, although the fact that the ratio of two geometrical magnitudes is not necessarily exactly representable by such numbers appears to have been discovered at a very early period. Euclid gave, in the fifth book of his treatise, a discussion of the theory of ratios, and in the tenth book, a theory of those incommensurable magnitudes which are ideally constructible by means of straight lines and circles. In later times*, the idea was current that, to the ratio of any two magnitudes of the same kind, there corresponds a definite number; and in fact Newton in his *Arithmetica Universalis* expressly defines a number as the ratio of any two quantities. Before the recent development of the arithmetical theories of irrational number, and to a considerable extent even later, a number has been regarded as the ratio of a segment of a straight line to a unit segment, and the conception of irrational number as the ratio of incommensurable segments has been accepted as a sufficient basis for the use of such numbers in Analysis.

In accordance with the doctrine that Mathematical Analysis must rest upon a purely arithmetical basis, the introduction of irrational numbers into Analysis must be made without an appeal to our intuition of extensive magnitude, but rather by an extension of the conception of Number, resting on a further development of the ideas which have been here discussed in connection with the theory of rational numbers. The necessity for this extension of the domain of Number arises not only on account of the inadequacy of rational numbers for application to ideally exact measurement, but also, as will be explained later in detail, because the theory of limits, which is an essential element in Analysis, is incapable of any rigorous formulation apart from a complete arithmetical theory of irrational numbers.

Before the recent establishment of the theory of irrational numbers, no completely adequate theory of Magnitude was in existence. This is not surprising, if we recognize the fact that the language requisite for a complete description of relations of magnitudes must be provided by a developed Arithmetic.

16. The successive extensions of the domain of Number, by the introduction of fractional and of negative numbers, were suggested by the desirability of so completing the domain that the operations of division and subtraction, which are not always possible in the more limited domain, might

* A good short account of the history of this subject will be found in the Article I. A 3, "Irrationalzahlen und Konvergenz unendlicher Prozesse," by A. Pringsheim, in the *Encyclopädie der Math. Wissenschaften*, vol. I. See also M. Cantor, *Geschichte der Math.*, vol. I.

always be so in the more extended one. In the aggregate of rational numbers, the operations of addition, subtraction, multiplication, and division are always possible operations; but it can be readily shewn that the inverse operation involved in determining a fractional power of a rational number is not, in general, a possible one.

As the simplest case of this impossibility of such operation, we may take the problem of finding the square root of a positive integer m which is not a square number. It can be shewn that such a number has no square root within the aggregate of rational numbers.

If possible* let m be the square of a rational fraction p/q in its lowest terms; thus $p^2 - mq^2 = 0$. There always exists a positive integer λ such that $\lambda^2 < m < (\lambda+1)^2$; we then have $\lambda q < p < (\lambda+1)q$.

Now let us consider the identity

$$(mq - \lambda p)^2 - m(p - \lambda q)^2 = (\lambda^2 - m)(p^2 - mq^2) = 0.$$

From this identity it follows that m is the square of the rational number $(mq - \lambda p)/(p - \lambda q)$, of which the denominator is less than q, and this is contrary to the hypothesis that m is the square of the fraction p/q which is in its lowest terms. It thus appears that there exists no rational number of which the square is m.

On the formal side of Arithmetic, a demand for the extension of the domain of number arises from the impossibility of carrying out, with the requisite generality, certain operations, as in the example given above. Such extensions of the domain of number as are made when fractional, negative, irrational, and complex numbers are successively adjoined to the original integral numbers, are made in accordance with a principle known as that of the permanence of forms, which was first indicated by Peacock†, and further developed by Hankel‡. This principle may be stated in the form that, in order to generalize the conception of number, the following four requisites must be satisfied:

(1) Every operation which is represented by a formal expression involving the unextended class of numbers, and which does not result in the representation of a number of the unextended class, must have a meaning assigned to it of such a character that the formal expression may be dealt with according to the same rules as would be applicable if the expression represented one of the unextended class of numbers.

(2) An extended definition of number must be given, such that a formal

* This proof is given by Dedekind in his tract *Stetigkeit und irrationale Zahlen.* An extension of Dedekind's method to the case of nth roots has been given by S. M. Jacob. See *Proc. Lond. Math. Soc.* Ser. 2, vol. I, p. 166.

† *British Association Report* for 1834; also *Symbolical Algebra*, Cambridge, 1845.

‡ See his *Theorie der komplexen Zahlsysteme*, Leipzig, 1867.

expression, as in (1), may represent a number in the extended sense of the term.

(3) A proof must be given, that for numbers of the extended class the same formal laws of operation hold as for the unextended class.

(4) Definitions must be given of the meaning of greater, equal, and less, in the extended domain of number, these terms being taken in the ordinal sense.

The arithmetical theory of irrational numbers has been developed in three main forms, of which the first* was given by Weierstrass in his lectures on Analytical Functions; the second† is that of G. Cantor, which was developed in further detail by Heine‡, and was also developed independently by Ch. Méray§; the third, that of R. Dedekind||, appeared about the same time as that of Cantor. We shall give an account of the theories of Dedekind and of Cantor, and shall shew that they are fundamentally identical.

KRONECKER'S SCHEME OF ARITHMETIZATION.

17. As it is now generally understood, the term "arithmetization" is used to denote the movement which has resulted in placing analysis on a basis free from all notions derived from the idea of measurable quantity, the fractional, negative, and irrational numbers being so defined that they depend ultimately upon the conception of integral number. An extreme theory of arithmetization has however been advocated by Kronecker¶, who proposed the abolition of all modifications and extensions of the conception of number, the integral numbers being alone retained. His ideal** is that every theorem in analysis shall be stated as a relation between integral numbers only, the terminology involved in the use of negative, fractional, and

* For an account of this Theory see S. Pincherle, *Giorn. di mat.*, vol. XVIII (1880), p. 185; also O. Biermann, *Theorie der analytischen Funktionen*, Leipzig, 1887, p. 19.

† *Math. Annalen*, vol. V (1872); see also *Math. Annalen*, vol. XXI, where Cantor discusses all the three theories.

‡ Crelle's *Journal*, vol. LXXIV (1872).

§ *Nouveau Précis d'Analyse infinitésimale*, Paris, 1872.

|| *Stetigkeit und irrationale Zahlen*, Brunswick, 1872.

¶ See Crelle's *Journal*, vol. CI, "Ueber den Zahlbegriff."

** He writes (*loc. cit.* p. 338) "Und ich glaube auch, dass es dereinst gelingen wird, den gesammten Inhalt aller dieser mathematischen Disciplinen zu 'arithmetisiren,' d. h. einzig und allein auf den im engsten Sinne genommenen Zahlbegriff zu gründen, also die Modificationen und Erweiterungen dieses Begriffs (ich meine hier namentlich die Hinzunahme der irrationalen sowie der continuirlichen Grössen) wieder abzustreifen, welche zumeist durch die Anwendungen auf die Geometrie und Mechanik veranlasst worden sind." He proceeds to shew in detail, how the notions of negative, fractional, and algebraical numbers can be avoided by substituting for equalities in which these numbers occur, congruences relative to certain moduli or systems of moduli. A similar suggestion had been made by Cauchy with reference to imaginary numbers.

irrational numbers, being entirely removed. This ideal, if it were possible to attain it, would amount to a reversal of the actual historical course which the science has pursued; for all actual progress has depended upon successive generalizations of the notion of number, although these generalizations are now regarded as ultimately dependent on the whole number for their foundation. The abandonment of the inestimable advantages of the formal use in Analysis of the extensions of the notion of number could only be characterised as a species of Mathematical Nihilism.

THE DEDEKIND THEORY OF IRRATIONAL NUMBERS.

18. Let us consider the aggregate of all the rational numbers ordered in the manner which has been previously discussed, and let us take any one such number N. We may conceive all the rational numbers to be divided into two classes R_1, and R_2, such that every number of R_1 is, in the ordinal sense of the term, less than every number belonging to the second class R_2, the two classes being separated by the number N, which may itself be assigned at choice either to the first or to the second class. If N belongs to the first class, it is the greatest number in that class, and the numbers of the second class have no number which is less than all the others of that class; if N be taken to belong to the second class, it is the least number in that class, and there exists no number in the first class which is greater than all the others; for if any rational number less than N be taken, it is always possible to find another greater one which is less than N. Such a division of the rational numbers into two classes is called a *section* (Schnitt), and we therefore say that *corresponding to any given rational number there exists a section which divides the aggregate of rational numbers into two classes, such that all the numbers of the first class are less than all those of the second class; and such that either in the first class there is no greatest number, or else in the second class there is no least number.*

It can be shewn by means of examples, that sections of the aggregate of rational numbers exist which are different in character from those just described. If m is a positive integer which is not a square number, we may conceive the rational numbers to be divided into two classes, the first of which contains all the negative numbers and also those positive numbers of which the square is less than m, including zero; the second class contains all the positive numbers of which the square is greater than m. The first class contains no greatest number, and the second class contains no least number; this section is said to be related to an irrational number $\sqrt{m}$, in the same way as a section such as has been considered above is related to a rational number. This example shews that sections of the rational numbers R exist, such that R is divided into two classes R_1, R_2, where every number of R_1 is less than every number of R_2, and such that R_1 contains no number greater than all the others, and also R_2 contains no number less than all the others.

A new aggregate of objects, the *real numbers*, may now be defined as follows:

To every section (R_1, R_2) *of the aggregate* R *of rational numbers, such that every number of* R *belongs to one or other of the two classes* R_1, R_2, *and every number in* R_1 *is ordinally less than every number in* R_2, *there corresponds a real number.*

In case neither R_1 *contains a number which is ordinally greater than all the others in* R_1, *nor* R_2 *contains a number which is ordinally less than all the others in* R_2, *the real number corresponding to the section is said to be an irrational number.*

In case either R_1 *has a greatest number* x, *or* R_2 *has a least number* x, *the section is said to define a real number corresponding to the rational number* x.

The *real* number which corresponds to a *rational* number x, though conceptually distinct from x, has no properties distinct from those of x, and is usually denoted by the same symbol.

The definition of a real number can be put into a different and somewhat less abstract form, by employing the notion of a lower segment of the aggregate of rational numbers. A *lower segment of the aggregate* R *of rational numbers* is any class of rational numbers which contains no number greater than all the others, and such that if any number whatever of the class be taken, the class contains *all those numbers of* R which are less than that number. A lower segment of R is identical with one of Dedekind's classes R_1, in case R_1 contains no greatest number.

A real number may be defined to be a lower segment of the aggregate* R *of rational numbers; and thus every real number, whether irrational or not, is a definite class of rational numbers.*

In accordance with this definition, the *real* number 3, for example, is defined to be the aggregate of all rational numbers which are less than the *rational* number 3; the irrational number $\sqrt{3}$ is defined as the aggregate of all rational numbers which are either negative, or if positive have their squares less than 3, the number zero being also included in the aggregate.

The use, here adopted, of the term *real* number, is sanctioned by general usage. The employment of the term *real*, has originated from the contrasting of these numbers, not with rational numbers, but with complex numbers. The extension of the term Number to the real numbers, is justified by the fact that it is possible to define the operations of addition, multiplication, &c.,

* This form of the definition is that given by B. Russell, see *The Principles of Mathematics*, vol. I, chaps. xxiii and xxiv; it was suggested by Peano, see *Rivista di Matématica*, vol. VI, pp. 126—140.

for real numbers, so that the formal laws of these operations are in agreement with those which hold for operations within the domain of the rational numbers.

19. It will now be shewn that the aggregate of real numbers, defined in Dedekind's manner, can be so ordered, that every real number has a definite rank in the aggregate, *i.e.* of any two real numbers it is determinate which has the higher and which the lower rank.

The basis of the scheme of order being taken to be the ordered aggregate of rational numbers, let us denote by n, n' any two real numbers, and let the sections by which they are defined be denoted by (R_1, R_2), (R_1', R_2') respectively.

The following cases may arise:

(1) If (R_1, R_2) and (R_1', R_2') are identical, that is, if every number in R_1 is also in R_1', and every number in R_2 is also in R_2', the two numbers n, n' are identical; thus $n \equiv n'$.

(2) Let us next suppose that there is one rational number $r_1 \equiv r_2'$, which is contained in R_1, but not in R_1'; it is consequently contained in R_2'. All the numbers in R_1' are less than r_2', and hence all the numbers in R_1' are in R_1. Since r_1 is the only number in R_1 which is contained in R_2', it follows that r_1 is greater than all the other numbers in R_1; and thus the number n defined by (R_1, R_2) is a number corresponding to the rational number r_1 or r_2'. All the elements in R_1' are contained in R_1, and are less than r_2'; all the numbers in R_2' except r_2', are greater than r_2', for if not they would be contained in R_1: hence the section (R_1', R_2') defines the real number $n' \equiv n$, corresponding to the rational number $r_2' \equiv r_1$. The two sections are essentially identical, the only difference being that the rational number $r_1 \equiv r_2'$, is regarded as belonging to the first class in one section and to the second class in the other section.

(3) If there are two different numbers belonging to R_1 which also belong to R_2', there are an indefinite number of other numbers which have the same property, since an unlimited number of rational numbers can be found which lie between two given rational numbers. In this case we define the number n or (R_1, R_2), to be greater, in the ordinal sense of the term, than n' or (R_1', R_2'), agreeably with the definition already down for the rational numbers.

The cases in which one, or more than one number which belongs to R_1' also belongs to R_2, may be treated in a similar manner; thus we define the meaning of the relation $n < n'$. It is easily seen that if $n > n'$, and $n' > n''$, then the relation $n > n''$ is also satisfied. Thus the system of real numbers is arranged in a regular order, such that those of them which correspond to rational numbers have the same relative rank as the corresponding rational numbers have in the aggregate of rational numbers.

20. The aggregate of real numbers has the following properties:

(1) If $\alpha > \beta$, and $\beta > \gamma$, then $\alpha > \gamma$.

(2) Between any two real numbers α, γ there are an unlimited number of real numbers. This is easily proved from the corresponding property of rational numbers, by considering the sections which define the numbers.

(3) If α is a fixed real number, then all real numbers may be divided into two classes $\bar{R}_1$, $\bar{R}_2$, such that $\bar{R}_1$ contains all the real numbers which are less than α, and $\bar{R}_2$ contains all real numbers which are greater than α. The number α may be regarded either as belonging to $\bar{R}_1$, in which case it is the greatest number in $\bar{R}_1$, or else as belonging to $\bar{R}_2$, in which case it is the least number in $\bar{R}_2$. This also follows from the definition above.

(4) If the aggregate of real numbers falls into two classes $\bar{R}_1$, $\bar{R}_2$, such that every number of $\bar{R}_1$ is less than every number of $\bar{R}_2$, then there exists one and only one number by which this section is produced.

To prove this, we observe that the section $(\bar{R}_1, \bar{R}_2)$ of the aggregate of real numbers also defines a section (R_1, R_2) of the aggregate of rational numbers, such that all rational numbers belonging to R_1 correspond to real numbers which belong to $\bar{R}_1$, and all numbers belonging to R_2 correspond to real numbers which belong to $\bar{R}_2$.

Let N be the real number defined by the section (R_1, R_2), and let N' be any real number different from N, defined by the section (R_1', R_2'). There are an indefinite number of rational numbers n which belong to only one of the aggregates R_1, R_1'; let $\bar{n}$ be the real number corresponding to n. If $N' < N$, then n belongs to R_1, and therefore $\bar{n}$ belongs to $\bar{R}_1$; and since $N' < \bar{n}$, it follows that N' belongs to $\bar{R}_1$. Similarly, if $N' > N$, we can shew that N' belongs to $\bar{R}_2$. It has thus been shewn that every number different from N belongs to $\bar{R}_1$ or to $\bar{R}_2$, according as it is less or greater than N. Thus N is either the greatest number in $\bar{R}_1$ or the least in $\bar{R}_2$, and therefore N is the only number by which the section $(\bar{R}_1, \bar{R}_2)$ can be made.

21. The operations between two real numbers may, in accordance with the above definition of real numbers by means of sections, be so defined that the result of each operation corresponds to a section of the rational numbers; thus the arithmetical operations are reduced to operations with rational numbers.

A complete theory of the operations involving real numbers can be established; and the formal laws of the operations can be shewn to be the same as in the case of the rational numbers, the range of possibility of operations being greater in the case of real than in that of rational numbers. This theory has been worked out to some extent by Dedekind: but as the Cantor theory of real numbers lends itself to a simpler detailed treatment of the operations than that of Dedekind, and as it will appear that the two theories are fundamentally equivalent to one another, it will be sufficient, as an example of the general method of treating operations in accordance with Dedekind's theory, to take only the case of the addition of two real numbers.

Let a, b be two real numbers defined by means of the sections (R_1, R_2), (R_1', R_2') respectively; then the sum $a+b$, of a and b, is defined by means of a section (R_1'', R_2'') which satisfies the following conditions:—If c_1 is any rational number, it is put into the class R_1'', provided there are two rational numbers a_1 in R_1, and b_1 in R_1', such that $a_1 + b_1 \geqq c_1$; all rational numbers c_2 for which this is not the case fall into the class R_2''. It is clear that every number c_1 is less than every number c_2, hence the section (R_1'', R_2'') is defined by means of this condition.

It can be shewn that, when a, b both correspond to rational numbers, this definition is in agreement with the ordinary definition of the sum of two rational numbers, so that the sum of the numbers corresponds to the sum of the corresponding rational numbers. Every number c_1 in R_1'', is $\leqq a+b$, because $a_1 \leqq a$, $b_1 \leqq b$, and therefore $a_1 + b_1 \leqq a+b$. Further, if there were contained in R_2'' a number $c_2 < a+b$, so that $a+b = c_2 + p$, where p is a positive rational number, we should have $c_2 = (a - \frac{1}{2}p) + (b - \frac{1}{2}p)$, and this is contrary to the definition of c_2, because $a - \frac{1}{2}p$ belongs to R_1, and $b - \frac{1}{2}p$ to R_1'; thus every number c_2 in R_2'', is $\geqq a+b$, and it has consequently been shewn that (R_1'', R_2'') defines the number $a+b$. As is usual, we have denoted the rational numbers a, b and the conceptually distinct real numbers a, b by the same symbols.

THE CANTOR THEORY OF IRRATIONAL NUMBERS.

22. The Cantor theory of irrational numbers essentially depends upon the use of convergent simply infinite ascending aggregates, or convergent sequences (Fundamentalreihen) in which the elements are rational numbers; we therefore proceed to define and discuss these aggregates.

A simply infinite ascending aggregate $(a_1, a_2, a_3, \ldots a_n, \ldots)$ *in which each element is a rational number, is said to be convergent, if it is such that corresponding to any fixed arbitrarily chosen positive rational number* ϵ, *as small, in the ordinal sense, as we please, a number* n *can be found such that* $|a_n - a_{n+m}| < \epsilon$, *for* $m = 1, 2, 3, \ldots$.

The symbol $|x|$ is here used to denote that one of the two numbers $x, -x$, which is positive; $|x|$ is said to be the absolute value of x.

This definition is equivalent to the statement that, in a simply infinite convergent aggregate, an element can always be found whose absolute difference from any element whatever which comes after it is as small as we please.

It should be observed that the terms "as small as we please," or "arbitrarily small," as applied to a positive number which is at choice, have reference to the conception of order only, and not to the non-arithmetical notion of magnitude. These expressions denote only that the number can be so chosen as to be of lower rank than any other arbitrarily chosen positive number.

To each value of ϵ there corresponds a value of n, which will in general have to be increased when ϵ is made smaller.

We may denote the aggregate by the symbol $\{a_n\}$, and shall speak of it shortly as a convergent sequence; that it is simply infinite will in future be understood.

In a convergent sequence, corresponding to any arbitrarily chosen positive number ϵ, *a number* n *can be found such that from and after that value of* n *the absolute difference of any two elements is less than* ϵ.

For choose n so that $| a_n - a_{n+m} | < \frac{1}{2}\epsilon$, for all positive integral values of m; then $| a_{n+m} - a_{n+m'} | \leqq | a_n - a_{n+m} | + | a_n - a_{n+m'} | < \epsilon$.

In the convergent sequence $\{a_n\}$, if we choose n such that $| a_n - a_{n+m} | < \epsilon$, then for $m = 1, 2, 3, \ldots$, the value of a_{n+m} for all values of m, lies between $a_n + \epsilon$, and $a_n - \epsilon$; that is to say, from and after some value of n all the elements lie between two rational numbers whose difference is arbitrarily small. There exist therefore two positive numbers α, α', of which the smaller α' may be zero, such that, from and after some fixed value of n, all the elements lie in absolute value between α and α'.

23. *If the aggregate* $(a_1, a_2, \ldots a_n \ldots)$ *is such that, from and after some fixed element, each element is less than the following one, and if all the elements are less than some fixed number* N, *then the aggregate is a convergent sequence.*

For if the aggregate is not convergent, there must exist some positive number δ, such that an indefinite number of increasing values $n, n_1, n_2, \ldots$ of n can be found, for which $| a_n - a_{n_1} |, | a_{n_2} - a_{n_1} |, | a_{n_3} - a_{n_2} | \ldots$ are all $\geqq \delta$. Since $a_{n_1} - a_n, a_{n_2} - a_{n_1}, \ldots$ are all positive, we have $a_{n_r} \geqq a_n + r\delta$, where r can always be taken so large that $a_n + r\delta > N$, or $a_{n_r} > N$, which is contrary to the hypothesis. Hence the aggregate is convergent.

It may in a similar manner be shewn that the aggregate is convergent if, from and after some fixed element, each element is greater than the following one, and if all the elements are greater than some fixed number.

If $\{a_n\}$, $\{b_n\}$ *are two convergent sequences of rational numbers, a value of* n *can be found corresponding to any arbitrarily assigned number* ϵ, *such that both* $| a_{n+m} - a_{n+m'} |$ *and* $| b_{n+m} - b_{n+m'} |$ *are less than* ϵ, m *and* m' *having all positive values.*

For we have only to choose for n the greater of the two values corresponding to ϵ, for each aggregate separately.

24. It will now be shewn that *the aggregates*

$$\{a_n + b_n\},\ \{a_n - b_n\},\ \{a_n b_n\},\ \left\{\frac{a_n}{b_n}\right\},$$

in which the elements are the sum, difference, product, and quotient, respectively of the corresponding elements of the two convergent sequences $\{a_n\}$, $\{b_n\}$, *are also convergent sequences, with a certain restriction in the last case.*

We have $|(a_n \pm b_n) - (a_{n+m} \pm b_{n+m})| \leqq |a_n - a_{n+m}| + |b_n - b_{n+m}|$; now n can be so chosen for a given ϵ, that for all values of m, $|a_n - a_{n+m}| < \frac{1}{2}\epsilon$, and $|b_n - b_{n+m}| < \frac{1}{2}\epsilon$, hence so that $|(a_n \pm b_n) - (a_{n+m} \pm b_{n+m})| < \epsilon$; therefore the aggregates $\{a_n + b_n\}$, $\{a_n - b_n\}$ are convergent.

Again,

$$\begin{aligned}|a_n b_n - a_{n+m} b_{n+m}| &= |a_n(b_n - b_{n+m}) + b_{n+m}(a_n - a_{n+m})| \\ &< \alpha\,|b_n - b_{n+m}| + \beta\,|a_n - a_{n+m}|\end{aligned}$$

where α, β are the two positive numbers which are such that $|a_n| < \alpha$, $|b_{n+m}| < \beta$, for all values of n and m.

We can take n so large that $|b_n - b_{n+m}| < \delta$, $|a_n - a_{n+m}| < \delta$, where δ is at our choice, and may be taken to be $\frac{\epsilon}{\alpha+\beta}$. Hence for this value of n, $|a_n b_n - a_{n+m} b_{n+m}| < \epsilon$, for every value of m; and thus $\{a_n b_n\}$ has been shewn to be a convergent sequence.

Lastly, in the case of $\left\{\frac{a_n}{b_n}\right\}$, we shall suppose that all the elements of $\{b_n\}$ are numerically greater than some fixed positive number β'.

We have then

$$\frac{a_n}{b_n} - \frac{a_{n+m}}{b_{n+m}} = \frac{a_n(b_{n+m} - b_n) + b_n(a_n - a_{n+m})}{b_n b_{n+m}},$$

hence

$$\left|\frac{a_n}{b_n} - \frac{a_{n+m}}{b_{n+m}}\right| < \frac{\alpha\,|b_n - b_{n+m}| + \beta\,|a_n - a_{n+m}|}{\beta'^2}.$$

If now n be chosen so that $|b_n - b_{n+m}|$, $|a_n - a_{n+m}|$ are both less than $\frac{\beta'^2}{\alpha+\beta}\epsilon$, for every value of m, then, for such a value of n, $\left|\frac{a_n}{b_n} - \frac{a_{n+m}}{b_{n+m}}\right| < \epsilon$; therefore $\left\{\frac{a_n}{b_n}\right\}$ is a convergent sequence, provided $|b_n|$ is, for all values of n, greater than some fixed positive number β', which may be as small as we please, but must not be zero.

25. The essence of Cantor's theory consists in the postulating of the existence of an aggregate of objects for thought, the real numbers, ordered in a definite manner, which manner is assigned by means of certain prescribed rules. Any element of the aggregate of real numbers is regarded as capable of symbolical representation by means of a convergent sequence of which the elements are rational numbers; and the mode in which the aggregate of real numbers is ordered is specified by means of formal rules relating to these convergent sequences. The aggregate of real numbers contains within itself an aggregate of objects which is similar to the ordered aggregate of rational numbers which has already been considered, in the sense that to each rational number there corresponds a certain real number; and the relative order of any two rational numbers, in the ordered aggregate of rational numbers, is the same as the relative order of the two corresponding real numbers in the new aggregate of real numbers. The rational numbers are frequently regarded as identical with the real numbers to which they correspond, and are denoted by the same symbols. In the development of Analysis, this identity leads to no difficulties; but in the fundamental theory of the aggregate of real numbers, a conceptual distinction between rational numbers and the real numbers to which they correspond must be made, in order to obviate logical difficulties, and especially with a view to coordinating Cantor's theory with that of Dedekind. Those real numbers which do not correspond to rational numbers are called irrational numbers; and those real numbers which correspond to rational numbers are usually spoken of as themselves rational numbers.

The rules by which the order of the real numbers in their aggregate is assigned are the following:

(1) Any convergent sequence $\{a_n\}$, of which the elements are rational numbers, is taken to represent a real number, which we may denote by a. Two such aggregates $\{a_n\}$, $\{b_n\}$ are taken to represent the same real number provided they satisfy the condition that, for any arbitrarily chosen positive rational number ϵ, a value of n can be found such that $|a_{n+m} - b_{n+m}| < \epsilon$, for this value of n, and for all values 0, 1, 2, 3 ... of m. Symbolically*, we have $\{a_n\} \equiv \{b_n\}$ under the condition stated.

(2) The real number represented by $\{a_n\}$ is regarded as of higher rank, or in the ordinal sense greater, than the real number represented by $\{b_n\}$, if, corresponding to any arbitrarily chosen positive rational number ϵ, a value of n can be found such that $a_{n+m} - b_{n+m}$ is positive for this value of n, and for all values 0, 1, 2, 3, ... of m, and greater than some fixed positive rational number δ which may be dependent upon ϵ. If, under similar conditions, $a_{n+m} - b_{n+m}$ is negative and numerically greater than some fixed positive number δ, the number represented by $\{a_n\}$ is taken to be less than that represented by $\{b_n\}$.

* Those who hold the view, advocated by Heine and others (see § 6, note), that a real number is identical with the set of symbols by which it is represented, can attach no direct meaning to this equality. It can only be taken to indicate that the two expressions may be used indifferently in any operation which involves the number.

The aggregate $(x, x, x, \ldots)$ or $\{x\}$, in which all the elements are identical with one rational number x, represents, since it is a convergent sequence, a real number which corresponds to the rational number x. It is clear, from the definition of order in (2), that the relative order of any two rational numbers, in the aggregate of rational numbers, is the same as that of the real numbers which correspond to them, in the aggregate of real numbers. The aggregate of rational numbers, and that of the real numbers which correspond to them, are *similar* aggregates.

Cantor's theory of irrational numbers, in the form in which it was presented by himself and by Heine, has been criticized* on the ground that an assumption is made that the sequence $\{x\}$, in which all the elements are the same rational number x, represents the rational number x itself, and that this amounts to an assumption that x is the limit of the sequence $\{x\}$; whereas the theory of arithmetical limits is represented by Cantor† as deducible from his theory of irrational numbers, and as not assumed in the construction of the theory itself. The theory in the form presented above is not open to this objection.

It can be shewn that any two convergent sequences $\{a_n\}$, $\{b_n\}$, satisfy one or other of the conditions laid down in the above definitions of equality and inequality, *i.e.* symbolically $\{a_n\} \gtreqless \{b_n\}$.

For, as has been shewn in § 22, corresponding to any arbitrarily chosen positive rational number δ, a value of n can be found such that a_{n+m} lies between $a_n + \delta$, and $a_n - \delta$, and such that, for the same value of n, b_{n+m} lies between $b_n + \delta$, and $b_n - \delta$; from this it follows that, for such value of n, $a_{n+m} - b_{n+m}$ lies between $a_n - b_n + 2\delta$ and $a_n - b_n - 2\delta$; or $a_{n+m} - b_{n+m}$ differs from $a_n - b_n$ by not more than 2δ. If corresponding values of δ and n can be found, for which $a_n - b_n + 2\delta$, $a_n - b_n - 2\delta$ have the same sign, then $a_{n+m} - b_{n+m}$ has the same sign as $a_n - b_n$, and is numerically greater than a fixed number; the condition of inequality of $\{a_n\}$, $\{b_n\}$ is then satisfied. If no such values of δ and n can be found, then $a_{n+m} - b_{n+m}$ is numerically less than 4δ; and since δ is arbitrarily small, the condition of equality of $\{a_n\}$, $\{b_n\}$ is then satisfied.

Although Cantor's form of the theory of irrational numbers, or rather of real numbers, is more convenient for detailed development than is Dedekind's form, yet it lies under the disadvantage that the nature of any single real number is veiled by the fact that, although it is a unique object, it is capable of representation by an unlimited number of convergent sequences, and therefore that the formal character of the theory does not make it clear what such a number really is. The comparison between the

* See B. Russell, *The Principles of Mathematics*, vol. I, p. 285.

† See *Math. Annalen*, vol. XXI, p. 568.

two theories which* will be given later on will throw light upon this point: for it will be shewn that a convergent sequence of the rational numbers is sufficient to define a section, of the kind fundamental in Dedekind's theory; and this, as we have seen, is equivalent to the definition of a lower segment, which is itself a certain definite class of rational numbers.

26. *The sum* $a+b$, *of two real numbers represented by the sequences* $\{a_n\}$, $\{b_n\}$, *is defined to be the real number represented by the sequence* $\{a_n+b_n\}$; *and the difference* $a-b$ *is defined as the number represented by* $\{a_n-b_n\}$.

It has been shewn in § 24 that the two sequences $\{a_n+b_n\}$, $\{a_n-b_n\}$ are convergent.

If $\{a_n\}$, $\{b_n\}$ represent the same number, the sequence $\{a_n-b_n\}$ defines the real number zero; for the condition that $|a_{n+m}-b_{n+m}|<\epsilon$, where ϵ is arbitrarily small, for a sufficiently great value of n, and for $m=0, 1, 2, 3, \ldots$, is in this case satisfied.

The product ab, *of two real numbers, is defined to be the number represented by the sequence* $\{a_nb_n\}$, *which has been shewn in* § 24 *to be convergent.*

The quotient a/b *is defined to be the number represented by the convergent sequence* $\left\{\frac{a_n}{b_n}\right\}$.

The only restriction on this definition is that b is not to be zero; for, when this condition is satisfied, the elements of the sequence $\{b_n\}$ which represents b, can be so chosen as to satisfy the restrictive condition given in § 24, that $\left\{\frac{a_n}{b_n}\right\}$ may be convergent.

It is necessary to shew that the sum $a+b$, the difference $a-b$, the product ab, and the quotient $\frac{a}{b}$, of two numbers a, b, as they have been defined above, are definite numbers independent of the particular convergent sequences used to represent the numbers a and b. Thus it must be shewn that if $\{a_n\}=\{a_n'\}$, $\{b_n\}=\{b_n'\}$, then

$$\{a_n+b_n\}=\{a_n'+b_n'\},\quad \{a_n-b_n\}=\{a_n'-b_n'\},\quad \{a_nb_n\}=\{a_n'b_n'\},$$

and

$$\left\{\frac{a_n}{b_n}\right\}=\left\{\frac{a_n'}{b_n'}\right\}.$$

We have

$$|(a_{n+m}\pm b_{n+m})-(a'_{n+m}\pm b'_{n+m})| \leqq |a_{n+m}-a'_{n+m}|+|b_{n+m}-b'_{n+m}|.$$

Now n can be so chosen, corresponding to a fixed number ϵ, that

$$|a_{n+m}-a'_{n+m}|<\tfrac{1}{2}\epsilon,\ |b_{n+m}-b'_{n+m}|<\tfrac{1}{2}\epsilon,\ \text{for } m=0, 1, 2, 3, \ldots;$$

with this value of n, we now have the condition

$$|(a_{n+m}\pm b_{n+m})-(a'_{n+m}\pm b'_{n+m})|<\epsilon,$$

* In Tannery's work *Introduction à la théorie des fonctions d'une variable*, chap. I, the theory of irrationals is treated by a combination of the two methods of Cantor and Dedekind.

satisfied; and this is the condition that $\{a_n \pm b_n\}$ represents the same number as $\{a_n' \pm b_n'\}$.

Again,

$$|a_{n+m}b_{n+m} - a'_{n+m}b'_{n+m}| \leqq |a_{n+m}(b_{n+m} - b'_{n+m})| + |b'_{n+m}(a_{n+m} - a'_{n+m})|$$
$$< A\,|b_{n+m} - b'_{n+m}| + B\,|a_{n+m} - a'_{n+m}|$$

where A, B are fixed positive numbers. It is now clear that n may be so chosen that $|a_{n+m}b_{n+m} - a'_{n+m}b'_{n+m}| < \eta$, where η is an arbitrarily chosen positive number; thus $\{a_n b_n\}$, $\{a_n' b_n'\}$ represent the same number.

Again,

$$\left|\frac{a_{n+m}}{b_{n+m}} - \frac{a'_{n+m}}{b'_{n+m}}\right| \leqq \left|\frac{a_{n+m}}{b_{n+m}b'_{n+m}}(b_{n+m} - b'_{n+m})\right| + \left|\frac{b_{n+m}}{b_{n+m}b'_{n+m}}(a_{n+m} - a'_{n+m})\right|$$

whence it can easily be seen that the condition is satisfied that

$$\left\{\frac{a_{n+m}}{b_{n+m}}\right\} \text{ and } \left\{\frac{a'_{n+m}}{b'_{n+m}}\right\}$$

represent the same number.

It is readily seen that the same commutative, associative, and distributive laws hold for the operations between real numbers, as for those involving rational numbers.

27. *If, from and after some fixed element a_n, all the elements of $\{a_n\}$ are positive and greater than some fixed positive rational number δ, then the real number represented by $\{a_n\}$ is positive, i.e. it is ordinally greater than zero.*

For, if we take any convergent sequence $\{b_n\}$ which defines the number zero, we have $\{a_n\} > \{b_n\}$; because, for some fixed value of n, $a_{n+m} - b_{n+m}$ is certainly positive for all values of m, and is greater than a fixed positive number; since n can be taken so large that $a_{n+m} > \delta$, and $b_{n+m} < \delta'$, where δ' is a positive rational number chosen less than δ.

Similarly it may be shewn that *the number defined by $\{a_n\}$ is negative, if, from and after some fixed value of n, all the a_n are negative and numerically greater than some fixed rational number δ.*

The term "numerically greater" denotes that $|a_n| > |\delta|$, and thus refers to the absolute values of the numbers concerned.

It is easily seen, that *unless $\{a_n\}$ is such that, from and after some fixed value of n, all the elements have the same sign, then $\{a_n\}$ must represent the number zero.*

If $\{a_n\}$, $\{b_n\}$ define two different real numbers a, b, then there lie between a, b an unlimited number of those real numbers which correspond to rational numbers.

Suppose $a > b$, then there exist a rational positive number δ, and an integer n, such that for all positive integral values of m including zero,

$a_{n+m} - b_{n+m} > \delta$, $| a_n - a_{n+m} | < \epsilon$, $| b_n - b_{n+m} | < \epsilon$, where ϵ is a rational number chosen to be $< \frac{1}{2}\delta$. If we take any rational number x, which is $< \delta$, and $> \epsilon$, the number $\{a_n - x\}$, in which all the elements are identical, lies between $\{a_n\}$ and $\{b_n\}$, since, for every value of m, we have $a_{n+m} - (a_n - x) > x - \epsilon$; therefore a is greater than the real number which corresponds to $a_n - x$. Again, $(a_n - x) - b_{n+m} = (a_n - b_n) + (b_n - b_{n+m}) - x > \delta - \epsilon - x$, therefore provided x is chosen to be $< \delta - \epsilon$, the real number which corresponds to $a_n - x$ is greater than b, and thus lies between a and b. The rational number $a_n - x$ may be chosen in an unlimited number of ways, since x is any rational number whatever which lies between $\delta - \epsilon$ and ϵ.

CONVERGENT SEQUENCES OF REAL NUMBERS.

28. Convergent sequences will now be considered, of which the elements are real numbers. It might at first sight be imagined that we should be led, by the employment of such sequences, to a further extension of the domain of number; it will however be seen that this is not the case.

The definition of a convergent sequence of real numbers is precisely similar to the definition which has been given in the case of sequences of rational numbers; thus $(\alpha_1, \alpha_2, \ldots \alpha_n, \ldots)$ is a convergent sequence of real numbers, provided that, corresponding to each arbitrarily chosen positive real number η, a value of n can be found such that $| \alpha_n - \alpha_{n+m} | < \eta$, for $m = 1, 2, 3, \ldots$. If we conceive that each such convergent sequence of real numbers represents a single ideal object, and if we give definitions of equality and inequality, and of the fundamental operations, precisely analogous to those given in § 25 and § 26, and assume as before that a convergent sequence in which all the elements are identical with the real number α is taken to represent that one of the new aggregate of objects which corresponds to α, it will be shewn that the new aggregate of objects is similar to the aggregate of real numbers, *i.e.* to each of the new objects there corresponds one of the real numbers, and also that the relation of order between corresponding pairs of elements in the two aggregates is the same. It thus appears that the aggregate of new objects is practically identical with the aggregate of real numbers, since the two are ordinally similar. We saw however that there is no such relation between the aggregate of real numbers and that of rational numbers. Therefore the passage from rational numbers to real numbers involves a real extension of the domain of number; but the passage from real numbers to an aggregate of objects represented, in accordance with the rules referred to above, by convergent sequences of real numbers, does not lead to any essential extension of the domain of number.

Let $\{a_n\}$ be any convergent sequence of rational numbers, and let $\{\bar{a}_n\}$ denote the sequence of those real numbers which correspond to the rational numbers which form the elements of $\{a_n\}$. It can easily be shewn that $\{\bar{a}_n\}$

is a convergent sequence: for if ϵ is an arbitrarily chosen positive rational number, and $\bar{\epsilon}$ the corresponding real number, the condition of convergence of $\{a_n\}$ is that, for every ϵ, a value of n can be found such that a_{n+m} lies between $a_n+\epsilon$, $a_n-\epsilon$, for $m=1, 2, 3, \ldots$. It follows from this that $\bar{a}_{n+m}$ lies between $\bar{a}_n+\bar{\epsilon}$, $\bar{a}_n-\bar{\epsilon}$; and this ensures the convergence of the sequence $\{\bar{a}_n\}$. Conversely, we see that if $\{\bar{a}_n\}$ is convergent, so also is $\{a_n\}$.

Next, let $\{\alpha_n\}$ be a convergent sequence of real numbers; then, between α_n and α_{n+1}, a real number $\bar{a}_n$ can be found which corresponds to a rational number a_n. Let this be done for every pair of consecutive elements in $\{\alpha_n\}$, and let us consider the sequences $\{\bar{a}_n\}$, $\{a_n\}$.

Since $\bar{a}_n-\bar{a}_{n+m}=(\bar{a}_n-\alpha_n)+(\alpha_n-\alpha_{n+m})+(\alpha_{n+m}-\bar{a}_{n+m})$, we have

$$|\bar{a}_n-\bar{a}_{n+m}| \leqq |\bar{a}_n-\alpha_n|+|\alpha_n-\alpha_{n+m}|+|\alpha_{n+m}-\bar{a}_{n+m}|.$$

Now, corresponding to any real positive number δ, n may be so chosen that for every value of m, $|\bar{a}_n-\alpha_n|$, $|\alpha_n-\alpha_{n+m}|$, $|\alpha_{n+m}-\bar{a}_{n+m}|$ are each less than $\frac{1}{3}\delta$; hence for such a value of n

$$|\bar{a}_n-\bar{a}_{n+m}|<\delta, \text{ for } m=1, 2, 3, \ldots,$$

thus the sequence $\{\bar{a}_n\}$ is convergent.

Again, $\{\alpha_n\}-\{\bar{a}_n\}=\{\alpha_n-\bar{a}_n\}$, and $|\alpha_n-\bar{a}_n|<|\alpha_n-\alpha_{n+1}|$,

and, since $\{\alpha_n\}$ is convergent, n may be chosen so great that, for that and all higher values of n, all the differences $|\alpha_n-\alpha_{n+1}|$ are less than an arbitrarily fixed number, hence $|\alpha_n-\bar{a}_n|$ satisfies the same condition; and therefore the two convergent sequences $\{\alpha_n\}$, $\{\bar{a}_n\}$ satisfy the condition of equality, or they represent the same one of the new objects. It has been shewn above that since $\{\bar{a}_n\}$ is convergent, so also is $\{a_n\}$. Now $\{a_n\}$ corresponds to a single real number a; therefore to any convergent sequence $\{\alpha_n\}$, of which the elements are real numbers, there corresponds a real number a.

We have further to shew that, if $\{\alpha_n\}$, $\{\beta_n\}$ are two convergent sequences of real numbers, and a, b the corresponding real numbers as just determined, then $a \gtreqless b$, according as $\{\alpha_n\} \gtreqless \{\beta_n\}$.

We know that $a \gtreqless b$, according as $\{a_n\} \gtreqless \{b_n\}$, where $\{b_n\}$ denotes the sequence of rational numbers which defines b, in the same way as $\{a_n\}$ defines a. Now $\{\alpha_n\}-\{\beta_n\}=\{\alpha_n-\beta_n\}$, and $\alpha_n-\beta_n=(\alpha_n-\bar{a}_n)+(\bar{b}_n-\beta_n)+(\bar{a}_n-\bar{b}_n)$; and we can choose n so large that $|\alpha_n-\bar{a}_n|$ and $|\bar{b}_n-\beta_n|$ are each less than $\frac{1}{2}\eta$, where η is an arbitrarily chosen real positive number: therefore we see that $\alpha_n-\beta_n$ lies between $\bar{a}_n-\bar{b}_n+\eta$ and $\bar{a}_n-\bar{b}_n-\eta$. It follows easily that $\{\alpha_n\} \gtreqless \{\beta_n\}$, according as $\{\bar{a}_n\} \gtreqless \{\bar{b}_n\}$, or according as $\{a_n\} \gtreqless \{b_n\}$, and hence as shewn above, according as $a \gtreqless b$. It has now been shewn that the objects which are represented by convergent sequences of real numbers have the same ordinal relation to one another as the real numbers to which those sequences have been shewn to correspond.

It appears, from what has now been proved, that, *to every convergent sequence of real numbers there corresponds a real number which may be taken to be defined by means of that sequence.*

There does not necessarily exist any rational number which corresponds in the same sense to a convergent sequence of rational numbers. The property of the aggregate of real numbers here stated embodies the characteristic difference between that aggregate and the aggregate of rational numbers; for the latter does not possess the corresponding property. It is this property of the aggregate of real numbers which makes it suitable to be the field of the real variable in the Theory of Functions.

THE ARITHMETICAL THEORY OF LIMITS.

29. *If $x_1, x_2, x_3, \ldots x_n, \ldots$ is a sequence of real numbers such that a number x exists which has the property that, corresponding to any arbitrarily chosen positive number ϵ, a value of n can be found such that $|x - x_n|$, $|x - x_{n+1}|$, $|x - x_{n+2}| \ldots$ are all less than ϵ, then the number x is said to be the limit of the sequence $x_1, x_2, \ldots x_n, \ldots$. This fact may be denoted by the equation $x = \underset{n=\infty}{L} x_n$.*

This definition is known as the arithmetical definition of a limit, and was first given *, in a form substantially identical with the above, by John Wallis.

It will be observed that the above definition contains no assertion as to the necessary existence of a limit of a sequence of numbers, but contains only a statement as to the relation of the limit to the numbers of the sequence, in case that limit exists.

There cannot be two numbers which both satisfy the condition of being a limit of the same sequence. For, if possible, let x, x' be two such numbers and let $|x - x'| = \delta$. Choose a value of ϵ, less than $\frac{1}{2}\delta$; then numbers n, n' can be found such that $|x - x_{n+m}|$, $|x' - x_{n'+m}|$ for all values 0, 1, 2, 3, ... of m, are less than ϵ. Suppose $n > n'$, then $|x - x_n|$ and $|x' - x_n|$ are both less than ϵ, hence $|x - x'| < 2\epsilon < \delta$, which is contrary to the condition $|x - x'| = \delta$.

It will now be shewn that, if the numbers of the sequence $\{x_n\}$ are real numbers, and if the sequence is a convergent one, then the real number x defined in the manner explained in § 28, by the sequence $\{x_n\}$, is the limit of the sequence.

For the two sequences $\{x_n\}$, $\{x\}$ both define the same number x, and therefore satisfy the condition of equality, which is that $|x - x_{n+m}| < \epsilon$, for any arbitrarily chosen ϵ, provided n be sufficiently great, and this is the condition that x should be the limit of the sequence $\{x_n\}$. A sequence of real numbers which has a limit must be convergent. For if x is the limit of $\{x_n\}$,

* *Arithmetica Infinitorum* (1655), Prop. 43, Lemma. See M. Cantor's *Geschichte der Mathematik*, vol. II, p. 823.

then for a sufficiently large value of n, $|x - x_n|$, $|x - x_{n+1}|$, ... $|x - x_{n+m}|$, ... are all less than $\frac{1}{2}\epsilon$, where ϵ is arbitrarily chosen; now $|x_n - x_{n+m}| \leqq |x - x_n| + |x - x_{n+m}|$; hence $|x_n - x_{n+m}| < \epsilon$, which is the condition of convergence of $\{x_n\}$.

As the complete result we have now the theorem known as the General Principle of Convergence*:

The necessary and sufficient condition that a sequence $x_1, x_2, \ldots x_n, \ldots$ *of real numbers may have a limit, is that, corresponding to every arbitrarily chosen positive number* ϵ, *a value of* n *can be found such that* $x_n - x_{n+1}$, $x_n - x_{n+2}$, $x_n - x_{n+3}$, ... *shall be all numerically less than* ϵ.

This theorem, which contains the criterion for the existence of a limit as defined in accordance with the arithmetical definition of a limit, is a deduction from Cantor's theory of real numbers.

30. If the numbers of a sequence $\{x_n\}$ are rational numbers, instead of real numbers, the definition of the limit is applicable, and it is a necessary but not a sufficient condition for the existence of the limit, that the sequence should be convergent. Strictly speaking, if a convergent sequence of rational numbers has a limit, that limit is also a rational number; but from the existence of convergent sequences of rational numbers which have no limit there arises the necessity for the extension of the domain of number, so that in the extended domain every convergent sequence may have a limit; this extension has been carried out by substituting Real Number for Rational Number. However, although a convergent sequence of rational numbers which has no rational limit, has in this strict sense no limit at all, by reason of the convergent sequence of those real numbers which correspond to the rational numbers having an irrational number as limit, and since, as has been seen above, these real numbers are for practical purposes not distinguished from the rational numbers to which they correspond, it is usual to consider this irrational number to be the limit of the sequence of rational numbers. We may thus assert that any convergent sequence of rational numbers which has not a rational number as limit, has an irrational number as its limit. This assertion is a correct one for the practical purposes of Mathematical Analysis.

31. The method of limits, which is essential both to pure Analysis and to the applications of Analysis in Geometry and in Kinetics, had a geometrical origin in the Method of Exhaustions, which was applied by the Greek geometers to determine lengths, areas, and volumes, in simple cases. This method, supplemented by the notion of the numerically *infinite*, was developed in later times, in various forms, into a general method which formed

* This term "das allgemeine Convergenzprinzip" is due to P. Du Bois-Reymond; see his *Allgemeine Functionentheorie*.

the basis of the Infinitesimal Calculus. The traditional geometrical conception of a limit may be exemplified by the case of the determination of the length of a curve as the limit of a sequence of properly chosen inscribed polygons. The lengths of the perimeters of the polygons are regarded as continually approaching the required length of the curve whilst the number of sides of the polygons is continually increased. The limit, the length of the curve, is then regarded as actually reached at the end of a process described as making the number of sides of the polygon infinite, this mode of attainment of the limit being however inaccessible to the sensuous imagination, and disguising an actual qualitative change of a geometrical figure which possesses corners and is bounded by segments of straight lines, into one which has no corners and has a curvilinear boundary. No doubt was felt as to the existence of the limit, which was regarded as obvious from geometrical intuition. That a curve possesses a length, or an area, was considered to require no proof. The first mathematician who recognized the necessity for a proof of the existence of a limit was Cauchy, who gave a proof of the existence of the integral of a continuous function. That the logical basis of the traditional method of limits is defective has in recent times received *a posteriori* confirmation by the exhibition of continuous functions which possess no differential coefficient, and by many other cases of exception to what were regarded as ordinary results of analysis resting on the method of limits, which have been brought to light by those mathematicians who have been engaged in examining the foundations of analysis.

The arithmetical theory of limits, which is summed up in the general principle of Convergence, provides a definite criterion for the existence of the limit of a sequence of numbers; and a considerable part of modern analysis is concerned with obtaining special forms of the general criterion adapted for use in special classes of cases. The theory is essentially dependent upon the theory of irrational numbers; for, in default of an arithmetical theory of irrational numbers, all attempts to prove* the existence of a limit of a convergent sequence are doomed to inevitable failure; and this for the simple reason that a convergent sequence of rational numbers does not necessarily possess a limit which is within the domain of such numbers. The definition of real numbers by means of convergent sequences of rational numbers is not a mere postulation of the existence of limits to such sequences; it involves rather the introduction of an enlarged conception of number, of such a character that the scheme of ordered real numbers should form a consistent whole, and such that every convergent sequence of numbers in the domain of real number necessarily has a limit within that domain. The postulation of the existence of the aggregate of real numbers is justified by shewing that

* An interesting discussion of various methods which have been suggested of proving the existence of a limit will be found in Du Bois-Reymond's *Allgemeine Functionentheorie*.

a complete scheme of definitions and postulates can be set up for the elements of this aggregate, and that such a scheme does not lead to contradiction*. As regards the existence of limits in the case of lengths, areas, volumes, &c., referred to above, the order of procedure is a reversal of the traditional one, the existence of the limit being no longer inferred from geometrical intuition. For example, in the case of the determination of the length of a curve, that length is not assumed to be independently known to exist, but is defined as the arithmetical limit of the sequence of numbers which represent the perimeters of a suitable sequence of inscribed polygons. When this sequence is convergent, and its limit is independent of the particular choice of the polygons, subject to suitable restrictions, then the limit so obtained determines the length of the curve. In case no such limit exists, the curve is regarded as not having a length.

EQUIVALENCE OF THE DEFINITIONS OF DEDEKIND AND CANTOR.

32. In order to establish the equivalence of the definitions of irrational numbers, as given by Dedekind and by Cantor, it must be shewn that every convergent sequence of rational numbers defines uniquely a section of all the rational numbers, and that this section is the same for all convergent sequences which represent the same real number in accordance with rule (1) in § 25. Conversely, it must be shewn that any number defined by a section can also be represented by a convergent sequence of rational numbers.

To shew that, corresponding to the convergent sequence $\{x_n\}$ which, in accordance with the Cantor theory, defines the real number x, a section can be found: Let r be any rational number, and let $\bar{r}$ be the corresponding real number represented by $\{r\}$. The number $x - \bar{r}$ is represented by $\{x_n - r\}$; and if this number is not zero, then (see § 27), from and after some fixed value of n, $x_n - r$ has a fixed sign, positive or negative according to the value of r. A section of the rational numbers may now be defined as follows:—Let every number r such that $x_n - r$ is negative, from and after some fixed value of n, be placed in the class R_2; and let every number for which $x_n - r$ is positive, from and after some fixed value of n, be placed in the class R_1. If there exists a rational number r, such that neither of these cases arises, then $x \equiv \bar{r}$, and r may be put into either of the classes R_1, R_2. It has thus been shewn that a section of the rational numbers can be determined, corresponding to the convergent sequence $\{x_n\}$.

Next, let $\{x_n'\}$ be any other convergent sequence which represents the same real number x, as $\{x_n\}$ does. We have to shew that the section of the rational numbers which corresponds to $\{x_n'\}$, is identical with that which

* On this mode of regarding the aggregate of real numbers as dependent upon a complete consistent scheme of definitions and axioms, see Hilbert, "Ueber den Zahlbegriff," *Jahresber. d. deutsch. math. Vereinigung*, vol. VIII (1900).

corresponds to $\{x_n\}$. If, as before, r denote any rational number, we have $\{x_n - r\} \equiv \{x_n' - r\}$. Now a value of n can be found, from and after which, $x_n - r$ and $x_n' - r$ both have fixed signs independent of n, and they must have the same sign. It follows that a number r which belongs to the class R_1, must also belong to the class R_1', by which the section corresponding to $\{x_n'\}$ is defined; and also a number r which belongs to the class R_2, necessarily belongs to R_2', except in the case $\{x_n\} = \{x_n'\} = \bar{r}$. It has thus been shewn that the section (R_1, R_2) which corresponds to $\{x_n\}$, is identical with the section (R_1', R_2') which corresponds to $\{x_n'\}$.

33. To shew that a convergent sequence can always be found such as to define the number corresponding to a given section (R_1, R_2), we observe that two rational numbers can always be found, one of which is in R_1 and the other in R_2, and such that their difference is numerically less than a given arbitrarily small rational number ϵ. Let A be any rational number in R_1, and let ϵ' be a rational number $< \epsilon$. Then of the numbers $A + \epsilon', A + 2\epsilon', \dots A + r\epsilon', \dots$ there must be a last one $A + r\epsilon'$ which falls in R_1, for $A + n\epsilon'$ may be made as large as we please by taking n large enough; the next number $A + (r+1)\epsilon'$ is then in R_2; and these numbers $A + r\epsilon'$, $A + (r+1)\epsilon'$, whose difference is $\epsilon' < \epsilon$ are the two numbers required. Moreover, if B is a rational number in R_2, the two numbers may be so chosen that both lie between A and B; for we need only take ϵ' to be of the form $\frac{1}{s}(B - A)$, where s is a positive integer so chosen that $\frac{1}{s}(B - A) < \epsilon$.

Now let $\{\epsilon_n\}$ be any convergent aggregate of rational numbers, which has zero for its limit. Choose x_1 in R_1, and x_2 in R_2, so that $x_2 - x_1 < \epsilon_1$; next take x_3 in R_1, and x_4 in R_2, so that $|x_3 - x_4| < \epsilon_2$; and that x_3, x_4 both lie between x_1 and x_2. Proceeding in this way, we can take x_{2n-1}, x_{2n} rational numbers of different classes, so that $|x_{2n-1} - x_{2n}| < \epsilon_n$; then either of the sequences $\{x_1, x_3, x_5, \dots\}$, $\{x_2, x_4, \dots\}$, defines the number which is represented by the section (R_1, R_2).

To prove this, we observe that $\{x_{2n-1}\}$ is a convergent sequence, since all the elements are $< x_2$, and $x_1 < x_3 < x_5 \dots$.

Again, suppose a is a rational number belonging to R_2, we can shew that, provided a rational number b exists in R_2 which is less than a, then a is greater than all the numbers $x_1, x_3, \dots$ by more than $a - b$. For

$$a - x_{2n-1} = (a - b) + (b - x_{2n-1}) > a - b,$$

however small $b - x_{2n-1}$ may become. Hence, unless a is the smallest rational number in R_2, the real number $\{a\}$ which corresponds to a, is greater than the number $(x_1, x_3, \dots)$.

Again, the sequences $\{x_{2n-1}\}$, $\{x_{2n}\}$ represent the same number, since their difference is the aggregate $\{\epsilon_n\}$ which defines zero. It now appears, by reasoning similar to the above, that any number a in R_1 is such that the real number $\{a\}$ is less than the number $\{x_{2n}\}$, unless a is the greatest rational number in R_1.

If either R_1 has a greatest rational number, or R_2 has a least one, the real number $\{a\}$ which corresponds to this rational number a, is itself defined by (R_1, R_2), and is the number represented by either of the sequences $\{x_{2n-1}\}$, $\{x_{2n}\}$. In any case, either of these two sequences defines the number given by the section (R_1, R_2).

The complete equivalence of the two theories of Dedekind and of Cantor has now been established. The first theory operates with the whole aggregate of rational numbers, the second with sequences selected out of that aggregate.

THE NON-EXISTENCE OF INFINITESIMALS.

34. It should be remarked that, in assuming that every section of the aggregate of real numbers defines a single real number, it has been implicitly assumed that *if a, b are any two positive real numbers, such that $a < b$, then a positive integer n can be found such that $na > b$.*

This is the arithmetical analogue of the so-called principle of Archimedes.

If any real numbers existed which are ordinally greater than all the numbers $a, 2a, 3a, \ldots$, then a section of the aggregate of real numbers would be defined by considering all numbers greater than all the numbers $a, 2a, 3a, \ldots$ to be in one class, and all the remaining real numbers to be in the other class; and this section would define a real number N. If now ϵ be an arbitrarily chosen positive number less than a, then $N-\epsilon$ is a number which is less than some of the numbers $a, 2a, 3a, \ldots$; and there must be a first of this set of numbers such that $N-\epsilon$ is less than it. Let this be pa; thus $N-\epsilon < pa$, hence $N < pa+\epsilon < (p+1)a$; which is contrary to the hypothesis that no number na is in the class of numbers which are $> N$.

The property of the aggregate of real numbers which has been established may be denoted by the statement that *the aggregate of real numbers forms an Archimedean system*; and this property of the aggregate is essentially equivalent to the property that every section of the aggregate defines a single number of the aggregate.

A consequence of the fact that the aggregate of real numbers forms an Archimedean system, is that so-called infinitesimal numbers do not exist within the aggregate. Every positive number ϵ, being such that an integer n can be found such that $n\epsilon > 1$, is a finite number, in the sense in which finite numbers were distinguished from infinitesimals, in the older forms of the Infinitesimal Calculus. *In Arithmetical Analysis, the conception of the* *actually infinitesimal has no place.* When the expression "infinitesimal" is used at all, it is to describe the process by which a variable to which the numbers of a sequence converging to zero are successively ascribed, as values, approaches the limit zero; thus an infinitesimal is a variable in a state of flux, never a number. Such a form of expression, appealing as it does to a mode of thinking which is essentially non-arithmetical, is better avoided.

THE THEORY OF INDICES.

35. When m is a positive integer, and x a rational number, x^m was defined to denote $x \times x \times x \ldots \times x$ (m factors); and this definition may be extended to the case in which x is any number defined by a convergent sequence; so that if x is defined by $\{x_n\}$, x^m is defined by $\{x_n{}^m\}$. It thus appears that for any real numbers x, we have, provided m and n are positive integers, $x^m \times x^n = x^{m+n}$.

If we assume x^0 and x^{-m} to be defined as having such a meaning that this law of indices holds when m or n is zero, or a negative integer, we can at once interpret x^0 and x^{-m}; for

$$x^0 \times x^n = x^{n+0} = x^n, \text{ thus } x^0 = 1,$$

and

$$x^{-n} \times x^n = x^0 = 1, \text{ thus } x^{-n} = \frac{1}{x^n}.$$

When p/q is a rational fraction, we shall define $x^{\frac{p}{q}}$ to have such a meaning that the above law of indices holds when either or both of m, n may be rational fractions. With this assumption

$$x^{\frac{p}{q}} \times x^{\frac{p}{q}} \times x^{\frac{p}{q}} \ldots \times x^{\frac{p}{q}} \text{ (q factors)} = x^p;$$

hence $(x^{\frac{p}{q}})^q = x^p$; or $x^{\frac{p}{q}}$ is, if it exists, a number whose qth power is x^p. The problem of determining, if possible, a number $x^{\frac{p}{q}}$, is that of finding a number whose qth power is a given number; and it has been already shewn that this is not always a possible operation within the domain of rational number.

It will now be shewn that, in the domain of real numbers, *the operation of finding $x^{\frac{p}{q}}$ is always a possible one when x is positive, and also when x is negative; provided however that in this latter case, q is an odd number, or if it is even, p is not odd.*

The following lemma will be required:—*If α is any real positive number less than unity, a positive integer m can be found such that $\alpha^m < \epsilon$, where ϵ is an arbitrarily prescribed positive number, or, in other words, $\underset{n=\infty}{L}\, \alpha^n = 0$.*

Since $\alpha^n > \alpha^{n+1}$, the sequence $(\alpha, \alpha^2, \ldots \alpha^n, \ldots)$ is convergent.

Suppose, if possible, that the sequence represents a positive number k different from zero; then m may be so chosen that $\alpha^m, \alpha^{m+1}, \ldots$ all differ from k by less than the arbitrarily prescribed number δ, say $\alpha^m = k + \eta$, where $\eta < \delta$. We have therefore $\alpha^{m+1} = (k+\eta)\alpha < (k+\delta)\alpha$; now δ can be chosen to be equal to $\frac{k(1-\alpha)}{1+\alpha}$, then $\alpha^{m+1} < k - \delta$; and this is contrary to the condition imposed in the choice of m.

It follows that k cannot be different from zero; and thus the lemma is established.

Suppose now that a is any positive number, rational or not, which lies between N^q and $(N+1)^q$, where N is a positive integer; we shall first shew that a number $N+h$, where $h<1$, can always be found such that $a-(N+h)^q$ is positive, and less than $a - N^q$. We find by division

$$(N+h)^q - N^q = \{(N+h) - N\}\{(N+h)^{q-1} + (N+h)^{q-2}N + \ldots + N^{q-1}\};$$

hence, if h is positive and less than unity, $(N+h)^q - N^q$ lies between

$$qhN^{q-1} \text{ and } qh(N+1)^{q-1}.$$

Since $$a-(N+h)^q = (a-N^q) - \{(N+h)^q - N^q\},$$

we must take h not greater than $\frac{a-N^q}{q(N+1)^{q-1}}$, in order that $a-(N+h)^q$ may certainly be positive; and the difference $a-(N+h)^q$ is then less than

$$(a-N^q) - qhN^{q-1}.$$

Let $$h = \frac{a-N^q}{q(N+1)^{q-1}},$$

then $$a-(N+h)^q < (a-N^q)\left\{1 - \left(\frac{N}{N+1}\right)^{q-1}\right\}.$$

Let $N_1 = N+h$, then N_1 is such that

$$a - N_1^q < (a-N^q)\left\{1 - \left(\frac{N}{N+1}\right)^{q-1}\right\},$$

and $$N_1 > N.$$

In a similar manner, we can shew that a number N_2 exists which is $> N_1$, and such that

$$a - N_2^q < (a-N_1^q)\left\{1 - \left(\frac{N_1}{N_1+1}\right)^{q-1}\right\}.$$

Proceeding in this manner, we obtain a series of numbers $N, N_1, N_2, \ldots N_r \ldots$ such that $N_r > N_{r-1}$, and that $a - N_r^q$ is positive, and less than

$$(a-N_{r-1}^q)\left[1 - \left(\frac{N_{r-1}}{N_{r-1}+1}\right)^{q-1}\right].$$

We shall now shew that $(N, N_1, N_2, \ldots N_r \ldots)$ is a convergent sequence which defines a number whose qth power is a.

The sequence $\{N_r\}$ is convergent since $N_r > N_{r-1}$, and every N_r is less than $N+1$. The qth power of the number defined by this convergent sequence is $\{N_r^q\}$, and we shall shew that this defines the number a or $\{a\}$.

We have

$$a - N_r^q < (a - N^q)\left[1 - \left(\frac{N_{r-1}}{N_{r-1}+1}\right)^{q-1}\right]\left[1 - \left(\frac{N_{r-2}}{N_{r-2}+1}\right)^{q-1}\right]\cdots\left[1 - \left(\frac{N}{N+1}\right)^{q-1}\right]$$

$$< (a - N^q)\left[1 - \left(\frac{N}{N+1}\right)^{q-1}\right]^r$$

for
$$\frac{N}{N+1} < \frac{N_{r-s}}{N_{r-s}+1},$$

and hence
$$1 - \left(\frac{N}{N+1}\right)^{q-1} > 1 - \left(\frac{N_{r-s}}{N_{r-s}+1}\right)^{q-1}.$$

Now $1 - \left(\frac{N}{N+1}\right)^{q-1}$ is a proper fraction, hence from the lemma proved above, we infer that a power r of the expression can be found which is less than an arbitrarily chosen positive number, which number we may take to be $\frac{\epsilon}{a - N^q}$. Hence, corresponding to every ϵ, a number r can be found such that $a - N^q_{r+s} < \epsilon$, for $s = 0, 1, 2, \ldots$, and therefore the sequence $\{N_r^q\}$, defines the number $\{a\}$ or a.

If a is a positive proper fraction, we have $(a^2)^q < a$, hence we may take N to be equal to a^2, instead of to a positive integer. Then $a < (N+1)^q$; thus this value of N will play the same part as the integral value in the above proof, and the reasoning is the same as before.

36. It has now been shewn that in every case a real number can be found of which the qth power is a given positive number a. It thus appears that $x^{\frac{p}{q}}$ has an interpretation within the domain of real numbers, when x is any positive number, and $\frac{p}{q}$ is a positive rational fraction.

We interpret $x^{-\frac{p}{q}}$ to be such that

$$x^{-\frac{p}{q}} \times x^{\frac{p}{q}} = x^0 = 1,$$

or
$$x^{-\frac{p}{q}} = 1/x^{\frac{p}{q}}.$$

If x is a negative number $-x'$, we have $(-x')^{\frac{p}{q}}$, defined as a number whose qth power is $(-x')^p$; and $(-x')^p$ is x'^p or $-x'^p$, according as p is even or odd.

If p is even, $(-x')^{\frac{p}{q}}$ can be interpreted as the value of $x'^{\frac{p}{q}}$. If p is odd, and q is odd, $(-x')^{\frac{p}{q}}$ may be interpreted as $-x'^{\frac{p}{q}}$. When p is odd, and q is even, we have obtained no interpretation of $(-x')^{\frac{p}{q}}$.

To complete the theory of indices in such a way that $(-x')^{\frac{2r+1}{2s}}$ may have an interpretation, we should require a further extension of the conception of number. This further extension takes place by the introduction of complex number, which is however outside the limits imposed upon this work as a treatise dealing only with real number.

37. The only case in which x^n, for a positive x, has not been defined, is that in which n is not a rational number. To extend the definition to this case, we suppose n to be defined by a convergent sequence $\{n_r\}$, in which all the numbers n_r are rational. We shall shew that the aggregate $\{x^{n_r}\}$ is convergent, and the number which it defines we shall denote by x^n.

We have $x^{n_r} - x^{n_{r+s}} = x^{n_r}\{1 - x^{n_{r+s}-n_r}\}$; now, since $\{n_r\}$ is a convergent aggregate, all the numbers n_r are numerically less than some fixed number, and therefore $|x^{n_r}| < A$, where A is some fixed number.

First suppose $x > 1$, then

$$x^{n_r} - x^{n_{r+s}} = x^{n_{r+s}}(x^{n_r - n_{r+s}} - 1) = x^{n_r}(1 - x^{n_{r+s}-n_r}),$$

hence $$|x^{n_r} - x^{n_{r+s}}| < A\,|x^{|n_r - n_{r+s}|} - 1|.$$

Now let r be so chosen that, for all values of s,

$$|n_r - n_{r+s}| < \frac{1}{q},$$

where q is a positive integer; then

$$x^{|n_r - n_{r+s}|} - 1 < x^{\frac{1}{q}} - 1 < \frac{x-1}{1 + x^{\frac{1}{q}} + x^{\frac{2}{q}} + \ldots + x^{\frac{q-1}{q}}},$$

hence $$|x^{|n_r - n_{r+s}|} - 1| < \frac{x-1}{q},$$

or $$|x^{n_r} - x^{n_{r+s}}| < A\,\frac{x-1}{q},$$

and if q is chosen so that $\frac{1}{q} < \frac{\epsilon}{A(x-1)}$, where ϵ is a fixed number, we see that r may be so chosen that $|x^{n_r} - x^{n_{r+s}}| < \epsilon$, for all values of s, therefore $\{x^{n_r}\}$ is a convergent sequence.

If $x < 1$, then $\left\{\frac{1}{x^{n_r}}\right\}$ is a convergent sequence, and therefore $\{x^{n_r}\}$ is also convergent, since it is the quotient of $\{1\}$ and $\{x^{n_r}\}$. If $x = 1$, then $\{x^{n_r}\} = 1$. Thus in every case $\{x^{n_r}\}$ is a convergent sequence if $\{n_r\}$ is convergent.

Since $\{x^{n_r}\} \times \{x^{m_r}\} = \{x^{n_r+m_r}\}$, we see that the definition of x^n, when n is not rational, is such that the relation $x^m \times x^n = x^{m+n}$ is satisfied.

THE REPRESENTATION OF REAL NUMBERS.

38. The ordinary mode of representation of a real number is by means of a decimal, or more generally by a radix-fraction. When the decimal is non-terminating, this mode of representation is a case of the representation by a convergent sequence of rational numbers, in accordance with Cantor's theory. For example, the number π is represented by the sequence

$$(3,\ 3\cdot1,\ 3\cdot141,\ 3\cdot1415,\ 3\cdot14159,\ 3\cdot141592,\ \ldots),$$

where by known processes, any prescribed element can be found as the result of a definite number of arithmetical operations.

The general theorem will be established that *every positive real number N is uniquely representable by means of a non-terminating series of radix-fractions, of which r, the radix, is any integer* $\geqq 2$.

Of the numbers $0, r, 2r, 3r, \ldots$, there is (see § 34), of all those which are less than rN, a greatest one $c_0 r$, which may be zero; thus

$$rN > c_0 r, \text{ and } < (c_0+1)r;$$

it follows that
$$N = c_0 + \frac{N_1}{r},$$
where N_1 is a positive number less than r.

In a similar manner we obtain

$$N_1 = c_1 + \frac{N_2}{r},\quad N_2 = c_2 + \frac{N_3}{r},\ \ldots\ N_n = c_n + \frac{N_{n+1}}{r},$$

where $N_2, N_3, \ldots N_{n+1}$ are all $< r$; therefore

$$N = c_0 + \frac{c_1}{r} + \frac{c_2}{r^2} + \ldots + \frac{c_n}{r^n} + \frac{N_{n+1}}{r^{n+1}},$$

where $c_0, c_1, c_2, \ldots c_n$ are each of them positive integral or zero, and $0 < N_{n+1} < r$.

Since
$$N - \left(c_0 + \frac{c_1}{r} + \frac{c_2}{r^2} + \ldots + \frac{c_n}{r^n}\right) < \frac{1}{r^n},$$

and it has been shewn that $\frac{1}{r^n}$ has the limit zero as n is indefinitely increased, we see that the sequence, of which the nth element is

$$c_0 + \frac{c_1}{r} + \frac{c_2}{r^2} + \ldots + \frac{c_n}{r^n},$$

is convergent, and represents the real number N. This is expressed by

$$N = c_0 + \frac{c_1}{r} + \frac{c_2}{r^2} + \ldots + \frac{c_n}{r^n} + \ldots,$$

in which N is represented by a non-terminating radix-fraction.

Let us now consider the case in which N is a rational number $\frac{a}{b}$, in its lowest term. We have $a = \alpha_0 b + \beta_0$, where $\beta_0 < b$; and $r\beta_0 = \alpha_1 b + \beta_1$, where $\beta_1 < b$; $r\beta_1 = \alpha_2 b + \beta_2, \ldots, r\beta_{n-1} = \alpha_n b + \beta_n$, where $\beta_1, \beta_2, \ldots \beta_n$ are all less than b.

If one of the numbers β, say β_n, is zero, we have

$$N \equiv \frac{a}{b} = \alpha_0 + \frac{\alpha_1}{r} + \frac{\alpha_2}{r^2} + \ldots + \frac{\alpha_n}{r^n};$$

and thus N is expressed in terminating radix-fractions; this case can only arise when b contains only prime factors of r. The terminating series of radix-fractions can be replaced by a periodic one which does not terminate. For if we use $\alpha_n - 1$ instead of α_n, as the numerator of r^n, we have

$$r\beta_{n-1} = (\alpha_n - 1)\, b + b;$$

thus β_n becomes b instead of zero, and

$$rb = (r-1)\, b + b;$$

thus $\beta_n, \beta_{n+1}, \ldots$ are all equal to b; and $\alpha_{n+1}, \alpha_{n+2}, \ldots$ are all equal to $r - 1$. Thus N is represented by

$$N \equiv \frac{a}{b} = \alpha_0 + \frac{\alpha_1}{r} + \frac{\alpha_2}{r^2} + \ldots + \frac{\alpha_n - 1}{r^n} + \frac{r-1}{r^{n+1}} + \frac{r-1}{r^{n+2}} + \ldots$$

It thus appears that *a rational number, which in its lowest terms has a denominator which contains only prime factors of* r, *is capable of a double representation; (1) by a terminating series of radix-fractions; (2) by a non-terminating series of radix-fractions, of which the numerators after some fixed one are all* $r - 1$.

In case none of the numbers $\beta_1, \beta_2, \ldots \beta_n, \ldots$ vanishes, it is clear that since all these numbers are either $1, 2, 3, \ldots b - 1$, they cannot be all unequal. Suppose β_n is the first which is repeated, and let $\beta_n = \beta_{n+m}$; it is then clear that $\beta_{n+1} = \beta_{n+m+1}, \beta_{n+2} = \beta_{n+m+2}, \ldots$; and therefore the number is represented by a recurring series of radix-fractions.

39. When a number is defined by means of a convergent sequence of some special form, it is in general not immediately obvious whether the number is rational or irrational. Many special investigations relating to particular cases, and various general criteria have been given by well-known mathematicians.

One of the most important modes of such representation of a number is that by an endless continued fraction. This fraction may be regarded as an aggregate, each element of which is a finite continued fraction. Legendre established the fundamental theorem that a number represented by an endless continued fraction

$$\frac{a_1}{b_1 \pm}\, \frac{a_2}{b_2 \pm}\, \frac{a_3}{b_3 \pm} \cdots \frac{a_n}{b_n \pm} \cdots,$$

that is, by an aggregate of which the nth element is

$$\frac{a_1}{b_1 \pm}\frac{a_2}{b_2 \pm}\frac{a_3}{b_3 \pm}\cdots\frac{a_n}{b_n},$$

is irrational*, provided the positive integers a_n, b_n are such that for every value of n, $b_n - a_n \geqq 1$; except that when $b_n - a_n = 1$, for every value of $n \geqq m$, where m is some fixed number, and when at the same time the signs before all the fractions $\frac{a_n}{b_n}$, for $n > m$, are negative, then the continued fraction converges to unity, or to a rational fraction, according as $m = 1$, or $m > 1$.

This theorem contains as special cases the theorems previously established by Lambert, that e^x, $\tan x$, $\log_e x$, $\tan^{-1} x$, π are irrational for rational values of x. The irrationality of e and e^2 was first proved by Euler†. Legendre‡ himself applied the general theorem to prove the irrationality of π^2, although his proof was lacking in rigidity.

The following general theorem has been proved§ by Cantor:—

If b, b', b'', ... is a set of positive integers such that, q being any arbitrarily chosen integer, all the numbers 1, b, bb', $bb'b''$, ... from and after some fixed number of the sequence, are divisible by q; then any number N can be uniquely represented by

$$I + \frac{\lambda}{b} + \frac{\mu}{bb'} + \frac{\nu}{bb'b''} + \cdots,$$

where I is an integer, and λ, μ, ν, ... are integers such that

$$\lambda \leqq b - 1, \quad \mu \leqq b' - 1, \quad \nu \leqq b'' - 1.$$

Further, in order that the number N may be rational, it is necessary that, from and after some fixed term of the series, all the numbers λ, μ, ν, ... have their highest possible values. If this condition is not satisfied, N is irrational.

As an example of this theorem, the number e represented by

$$2 + \frac{1}{2} + \frac{1}{2 \cdot 3} + \frac{1}{2 \cdot 3 \cdot 4} + \cdots$$

is seen to be irrational.

A particular case of Cantor's theorem is that in which the sequence of numbers b, b', b'', ... from a particular element onwards, is periodic. In

* A proof of this theorem is given by Pringsheim, "Ueber die Convergenz unendlicher Kettenbrüche," *Sitzungsberichte d. bayer. Akad.* vol. XXVII, 1897, p. 318.

† On the history of these theorems see Pringsheim's article "Ueber die ersten Beweise der Irrationalität von e and π," *Sitzungsberichte d. bayer. Akad.* vol. XXVII.

‡ See his *Éléments de Géométrie*, Note 4; see also Rudio's work, "Archimedes, Huygens, Lambert, Legendre," 1892, p. 166.

§ Schlömilch's *Zeitschrift*, vol. XIV (1869), "Ueber die einfachen Zahlensysteme."

this case, the necessary and sufficient condition that the number represented by

$$\frac{\beta}{b}+\frac{\beta'}{bb'}+\frac{\beta''}{bb'b''}+\dots$$

should be rational, is that the sequence $\beta, \beta', \beta'', \dots$ be, from and after some fixed number of the sequence, periodic. This is a generalization of the theorem relating to a number represented by radix-fractions.

If $b=2,\ b'=3,\ b''=4, \dots$ we obtain the theorem* that the number represented by

$$\frac{c_1}{2!}+\frac{c_2}{3!}+\frac{c_3}{4!}+\dots+\frac{c_n}{n!}+\dots$$

where $c_n \leqq n-1$, is rational, only if, from and after some particular value of n, $c_n = n-1$.

A mode of representation of numbers by sequence of products has been given† by Cantor. He shews that every number $N>1$, can be uniquely represented in the form

$$\left(1+\frac{1}{a}\right)\left(1+\frac{1}{b}\right)\left(1+\frac{1}{c}\right)\left(1+\frac{1}{d}\right)\dots,$$

where $a, b, c, \dots$ are integers such that

$$b \geqq a^2,\ \ c \geqq b^2,\ \ d \geqq c^2,\ \dots.$$

The number a is determined as the integral part of $\frac{N}{N-1}$. If $\frac{Na}{a+1}=B$, b is the integral part of $\frac{B}{B-1}$; if $\frac{Bb}{b+1}=C$, c is the integral part of $\frac{C}{C-1}$; and so on.

As an example, $\sqrt{2}$ is represented by

$$\left(1+\frac{1}{3}\right)\left(1+\frac{1}{17}\right)\left(1+\frac{1}{577}\right)\left(1+\frac{1}{665857}\right)\dots,$$

where $\quad 17=2.3^2-1,\ \ 577=2.17^2-1,\ \ 665857=2.577^2-1, \dots.$

The criterion for determining whether N is rational or irrational is the following:—

The number represented by

$$\left(1+\frac{1}{a}\right)\left(1+\frac{1}{b}\right)\left(1+\frac{1}{c}\right)\dots,$$

where $\quad b \geqq a^2,\ \ c \geqq b^2, \dots,$

* See Stéphanos, *Bulletin de la soc. math. de France*, vol. VII (1879). For further information on the history of this subject see Pringsheim's article I. A. 3, in the *Encyclopädie der Math. Wissenschaften*.

† Schlömilch's *Zeitschrift*, vol. XIV (1869), "Ueber zwei Sätze...."

all the numbers $a, b, c, \ldots$ being positive integers, is rational if, from and after some fixed number of the sequence $a, b, c, \ldots$, each number is the square of the preceding number of the sequence; but the number is irrational if this condition is not satisfied.

THE CONTINUUM OF REAL NUMBERS.

40. If a_1, b_1 are any two real numbers such that $a_1 < b_1$, then two real numbers a_2, b_2, $(a_2 < b_2)$, can be found both lying between a_1, b_1, and such that the difference between a_2, b_2 is as small as we please, *i.e.* $b_2 - a_2 < \epsilon$, where ϵ is an arbitrarily prescribed number. Between a_2, b_2, two more numbers a_3, b_3, $(a_3 < b_3)$, can be found whose difference is again as small as we please; and this process may be carried on indefinitely. This property of the aggregate of real numbers may be expressed, to use the term introduced by G. Cantor, by saying that the aggregate of real numbers is *connex*; it arises from the fact that an indefinite series of numbers can be found which lie between any two given numbers. If we anticipate a term which will be introduced when we come to the general theory of aggregates, the property of connexity may be expressed by saying that *the aggregate of real numbers is everywhere-dense.*

It will further be observed that the aggregate of rational numbers is also connex, or everywhere-dense; so that, so far as this property is concerned, there is nothing to differentiate the one aggregate from the other.

If the difference of a_n and b_n is denoted by ϵ_n, and the sequence $\epsilon_1, \epsilon_2, \ldots \epsilon_n, \ldots$ satisfies the condition, that corresponding to any fixed arbitrarily small positive number η, a value of n can be found such that $\epsilon_n, \epsilon_{n+1}, \ldots$ are all less than η, then there exists a single real number x which is greater than all the numbers $a_1, a_2, \ldots$, and less than all the numbers $b_1, b_2, \ldots$. This number x is the limit of either of the sequences $(a_1, a_2, \ldots a_n, \ldots)$ and $(b_1, b_2, \ldots b_n, \ldots)$, and is defined by a section of all the real numbers.

If we confine ourselves to the domain of rational numbers, there subsists in that domain no such property; that is, the above numbers a, b being all rational, no such rational number as x necessarily exists.

In the domain of Real Number, every convergent sequence has a limit which is a number belonging to that domain; and, conversely, every number is the limit of properly chosen convergent sequences of numbers belonging to the domain: but in the domain of Rational Number a corresponding statement does not hold good, although the converse is still valid.

This property, which the domain of real numbers possesses, we express by saying that *the aggregate of real numbers is perfect. The aggregate of rational numbers is not perfect.*

From the point of view of Dedekind's theory, the property that the aggregate of real numbers is perfect expresses the fact that every section of the real numbers corresponds to a single real number, and the converse. A section of the rational numbers does not always correspond to a rational number; consequently the aggregate of rational numbers is not perfect.

We give the name *continuum** to an aggregate which possesses the two properties of being connex, and of being perfect. This is in the first instance taken to be the definition of the meaning of the word continuum, as it is used in Analysis. Thus the aggregate of real numbers forms a continuum; whereas the aggregate of rational numbers is essentially discrete, and does not form a continuum, since one of the two essential properties of a continuum is absent.

The aggregate of real numbers is spoken of as *the continuum of real numbers*, or *the arithmetic continuum.*

The real numbers which lie between two numbers a, b do not form a continuum, but if the two numbers a, b themselves are considered to be included in the total aggregate, then this completed aggregate does form a continuum.

All the real numbers x such that $a \leqq x \leqq b$, in the ordinal sense of the symbols $<$, $=$, $>$, are said to form an interval (a, b); and such an interval is frequently described as a closed interval.

The real numbers x which are such that $a < x < b$, are frequently said to form an *open* interval (a, b).

The closed interval (a, b) is a continuum, since it satisfies the two necessary conditions for the applicability of the term; but the open interval (a, b) is not a continuum, as it contains convergent sequences which have no limit belonging to the open interval. Such an open interval has been termed by Cantor a *semi-continuum.*

Of the two essential properties of the arithmetic continuum, that of *connexity*, and that denoted by the term *perfect*, the latter is absolutely indispensable, in order that the arithmetic continuum may be suitable to be the field of operations in analysis. It will appear, when we come to the consideration of the theory of functions of a real variable, that many of the most important properties of a function may still subsist even if the domain of the variable lacks the property of connexity; but that such properties would not belong to functions of a variable which is defined for a domain such that convergent sequences of numbers in it possess no limit within that domain, and which therefore lacks the property of being perfect. This is the more remarkable on account of the fact that, in the older traditional notion of a continuum, the property of connexity was the one which was regarded as all important; the more essential property of being

* See Cantor, *Math. Annalen*, vol. XXI, p. 576.

perfect has only been explicitly formulated in the course of construction of the modern arithmetical theory.

41. The term arithmetic continuum is used to denote the aggregate of real numbers, because it is held that the system of numbers of this aggregate is adequate for the complete analytical representation of what is known as continuous magnitude. The theory of the arithmetic continuum has been criticized on the ground that it is an attempt to find the continuous within the domain of number, whereas number is essentially discrete. Such an objection presupposes the existence of some independent conception of the continuum, with which that of the aggregate of real numbers can be compared. At the time when the theory of the arithmetic continuum was developed, the only conception of the continuum which was extant was that of the continuum as given by intuition; but this, as we shall shew, is too vague a conception to be fitted for an object of exact mathematical thought, until its character as a pure intuitional datum has been modified by exact definitions and axioms. The discussions connected with arithmetization have led to the construction of abstract theories* of measurable quantity; and these all involve the use of some system of arithmetic, as providing the necessary language for the description of the relations of magnitudes and quantities. It would thus appear to be highly probable that, whatever abstract conception of the intuitional continuum of quantity and magnitude may be developed, a parallel conception of the arithmetic continuum, though not necessarily identical with the one which we have discussed, will be required. To any such scheme of numbers, the same objection might be raised as has been referred to above; but if the objection were a valid one, the complete representation of continuous magnitudes by numbers would, under any theory of such magnitudes, be impossible. It is clear that it is only in connection with an exact abstract theory of magnitude, that any question as to the adequacy of the continuum of real numbers for the measurement of magnitudes can arise. For actual measurement of physical, or of spatial, or temporal magnitudes, the rational numbers are sufficient; such measurement being essentially of an approximate character only, the degree of error depending upon the accuracy of the instruments employed.

The purely ordinal nature of the conception of the arithmetic continuum, including the ordinal character of an interval, has been pointed out in the course of the development of the theory. This will be further elucidated in connection with the abstract theory of order-types to be discussed in Chapter III.

* See O. H. Hölder, *Die Axiome der Quantität und die Lehre vom Mass*, Leipziger Berichte, vol. LIII (1901); also Veronese's work, *Fondamenti di Geometria*, 1891; and Bettazzi's work, *Teoria delle grandezze*, 1890.

THE CONTINUUM GIVEN BY INTUITION.

42. Before the development of analysis was made to rest upon a purely arithmetical basis, it was usually considered that the field of operations was the continuum given by our intuition of extensive magnitude, especially of spatial or temporal magnitude, and of the motion of bodies through space.

The intuitive idea of continuous motion implies that in order that a body may pass from one position A to another position B, it must pass through every intermediate position in its path. An attempt to answer the question, what is meant by *every* intermediate position, reveals the essential difficulties of this conception, and gives rise to a demand for an exact theoretical treatment of continuous magnitude.

The implication contained in the idea of continuous motion, shews that, between A and B, other positions A', B' exist, which the body must occupy at definite times; that between A', B', other such positions exist, and so on. The intuitive notion of the continuum, and that of continuous motion negate the idea that such a process of subdivision can be conceived of as having a definite termination. The view is prevalent that the intuitional notions of continuity and of continuous motion are fundamental and *sui generis*; and that they are incapable of being exhaustively described by a scheme of specification of positions. Nevertheless, the aspect of the continuum as a field of possible positions is the one which is accessible to Arithmetic Analysis, and with which alone Mathematical Analysis is directly concerned. That property of the intuitional continuum, which may be described as unlimited divisibility, is the only one that is immediately available for use in Mathematical thought; and this property is not sufficient for the purposes in view, until it has been supplemented by a system of axioms and definitions which shall suffice to provide a complete and exact description of the possible positions of points and other geometrical objects which can be determined in space. Such a scheme constitutes an abstract theory of spatial magnitude.

The exact theory of magnitude was developed to a considerable extent by Euclid; but not until recently, under the influence of the ideas of the arithmetical theory, has it been perfected in a form which exhibits the exact system of axioms and definitions necessary for a characterization of continuity, that is adequate for mathematical analysis. Besides the arithmetic theory of number, there exists at the present time a theory of magnitude which runs to a certain extent parallel with the former theory. Some mathematicians* still prefer to regard number as primarily representing the ratio of two magnitudes; but they nevertheless to a large extent employ the methods of arithmetical analysis.

* P. Du Bois Reymond in his *Allgemeine Functionentheorie* strongly advocates the view that linear magnitude forms the basis of the conception of Number. See also Stolz, *Allgemeine Arithmetik*, where both views of Number are developed. See also G. Ascoli, *Rend. Ist. Lomb.* (2) 28 (1895).

THE STRAIGHT LINE AS A CONTINUUM.

43. Although it is no part of the plan of the present work to enter fully into the general theory of Magnitude, it is necessary briefly to consider the case of those magnitudes which are segments of a straight line, that straight line which is the ideal object of geometry, and which is the ideal counterpart of the physical straight line of perception.

The length of the segment between two points A, B, of a straight line, is a particular case of a magnitude; and we shall take this conception as a datum, subject to a set of axioms* relating to the notions of congruency, and to the notions greater and less as applied to magnitudes.

We assume that any number of congruent segments OA, AB, BC, ... can be constructed on the straight line; and that any segment OA can be divided into any number of segments which are all equal to one another.

Any segment OA may be taken as the unit of length, so that its magnitude is represented by the number 1; its multiples OB, OC, ... are denoted by the numbers 2, 3, ... If each one of the segments OA, AB, BC, ... be divided into the same number q of equal parts, then, if P is a point of division, OP is denoted by a fractional number p/q, where p is the number of the sub-segments in OP. Thus when p, q are any positive integral numbers, p/q represents a definite magnitude OP, the unit magnitude OA having been fixed upon beforehand.

Further, the number p/q may also be regarded as representing the position of the point P itself. In order to represent points of the straight line on both sides of O, the convention is made, that points on one side of O shall be represented by positive numbers, and those on the other side by negative numbers; thus if P is on the right of O, and P' on the left of O, and if $OP = OP'$, the point P' is represented by the number $-p/q$. The length of any segment of the straight line, whose ends are points to which rational numbers have been assigned in the manner explained above, is the difference of the above two numbers. In this manner, we have a correspondence established between the aggregate of rational numbers and an aggregate of points on the straight line, the relation of order being conserved in the correspondence, so that the two aggregates are similar.

The set of points, thus represented by rational numbers, we may speak of as the rational points of the straight line; but it must be remembered that a definite origin O, and a definite unit of length OA, are supposed to have been fixed upon beforehand; and if these be altered, the set of rational points will in general be altered also.

* These axioms are discussed by O. Hölder, *Leipziger Berichte*, vol. LIII, 1901.

It has been assumed as an axiom that, if P_1Q_1 is any segment of the straight line, it may be divided into any number n_1 of equal parts: of these, if P_2Q_2 be taken as one, the same axiom asserts that P_2Q_2 may be similarly divided into any number, n_2, of equal parts, P_3Q_3 being one of the parts; and that this process may be repeated an unlimited number of times. The axiom is equivalent to an assumption that the straight line is capable of unlimited divisibility; and this, being a characteristic property of the intuitional linear continuum, must also hold for its ideal counterpart, the straight line which we are here considering.

We proceed to assume as another axiom that, P_1Q_1, P_2Q_2, P_3Q_3 ... being the segments constructed as above, there exists in the straight line one point X, and one only, which separates all the points $P_1, P_2, P_3, \ldots$ from all the points $Q_1, Q_2, Q_3, \ldots$. If Y be any point other than X, then points belonging to the sequence $P_1, P_2, P_3, \ldots$ and points belonging to the sequence $Q_1, Q_2, Q_3, \ldots$ can be found which are both on the same side of Y.

The point X may be regarded as the limit of either sequence of points; and the property corresponds to that property of the arithmetic continuum which is expressed by saying that it is perfect.

In accordance with this axiom there is one single point on the straight line which corresponds to any given real number; and this point, or the magnitude of the corresponding segment, may be represented by the real number.

This axiom has been stated by Dedekind, in a form corresponding to his definition of an irrational number:—that a section of the rational points, in which they are divided into two classes, is made by a single point.

Another form of the axiom is that known as the *Axiom of Archimedes**:—that if AB, $A'B'$ are any two segments of the straight line, of which AB is the smaller one, an integer n can always be found such that $n \cdot AB > A'B'$. As in the case of the arithmetic continuum, this is equivalent to the negation of the existence of infinitesimal segments of the straight line.

This axiom being assumed, there is a complete correspondence between the points of the straight line and the aggregate of real numbers. Thus the nature of the linear continuum, that is, so far as its possible parts, and the possible positions in it, are concerned, is completely represented and described by means of the arithmetic continuum, the axioms relating to the straight line having been so chosen that this may be the case. It will be observed that there is no real disparity between the rational points and the irrational points of the straight line; a point, which with one origin and one unit of length, is a rational point, may be an irrational point if another origin, or another unit of length, be chosen.

* The importance of the Axiom of Archimedes in this connection was pointed out and discussed by Stolz, *Math. Annalen*, vols. XXII and XXXIX.

44. The mode which has been adopted above, of establishing a complete correspondence between the aggregate of real numbers and the aggregate of points in a straight line, though the most convenient mode, is not the only possible one. All that is really necessary for the correspondence is that, in accordance with some systematic scheme, the points in the straight line shall be made to correspond with the numbers of the arithmetic continuum in such a way that the relation of order is conserved in the correspondence. It is not necessary that the difference of two numbers should represent the length of the segment of the straight line which is terminated by the points that correspond to the two numbers. The mode of correspondence given above is however the simplest one, and will therefore be adopted for the purpose of enabling us to use the language of geometry in analytical discussion.

In the case of space of two or of three dimensions, it will be assumed as axiomatic that one point of the space, and one only, corresponds to each pair or triplet of real numbers which represent Cartesian coordinates. This axiom may be considered as fundamental in the Cartesian system of analytical geometry.

The disputable idea that the theory here explained necessarily implies that a continuum is to be regarded as made up of points, which are elements not possessing magnitude, has frequently been a stumbling-block in the way of the acceptance of the view of the spatial continuum which has been indicated above. It has been held that, if space is to be regarded as made up of elements, these elements must themselves possess spatial character; and this view has given rise to various theories of infinitesimals or of indivisibles, as components of spatial magnitude. The most modern and complete theory of this kind has been developed by Veronese *, and is based upon a denial of the principle of Archimedes which has been already referred to. In Veronese's system, when a unit segment of a straight line has been chosen, there exist segments which are too large, and others that are too small, to be capable of representation by finite numbers; and these segments are respectively infinite, and infinitesimal, relatively to the unit segment chosen. Under this scheme, a section of the rational points, or a section of the points represented by real numbers, is made, not by a single point, but by an infinitesimal segment. Veronese has consequently introduced systems of infinite and of infinitesimal numbers, each of an unlimited number of orders, for the measurement of segments which, relatively to a given scale, are infinite or infinitesimal. From his point of view, the points on a straight line which represent the real numbers form only a relative continuum, *i.e.* one which is relative to the particular scale of measurement

* See his *Fondamenti di Geometria*, Pisa, 1894; a German translation by Schepp has been published in Leipzig.

employed; and he contemplates the conception of an absolute continuum, for the representation of which his series of sets of infinite and infinitesimal numbers are requisite. A segment, which in a given scale is finite, may be infinitesimal, or infinite of any order, when measured relatively to another scale.

The validity of Veronese's system has been criticized by Cantor and others, on the ground that the definitions contained in it, relating to equality and inequality, lead to contradiction; it is however unnecessary for our purpose to enter into the controversy on this point. The straight line of geometry is an ideal object of which any properties whatever may be postulated, provided that they satisfy the conditions, (1) that they form a valid scheme, *i.e.* one which does not lead to contradiction, and (2) that the object defined is such that it is not in contradiction with empirical straightness and linearity. There is no *à priori* objection to the existence of two or more such adequate conceptual systems, each self-consistent, even if they be incompatible with one another; but of such rival schemes the simplest will naturally be chosen for actual use. Assuming then the possibility of setting up a valid non-Archimedean system for the straight line, still the simpler system, in which the principle of Archimedes is assumed, is to be preferred, because it gives a simpler conception of the nature of the straight line, and is adequate for the purposes for which it was devised. The case of the non-Euclidean systems of geometry is an instance of the existence of valid geometrical schemes divergent from one another, which nevertheless all afford a sufficient representation of physical space-percepts.

An answer to the difficult question, in what sense the straight line, or a space of two or of three dimensions, admits of being regarded as an aggregate of points, can only be discussed after a full treatment of the nature and properties of infinite aggregates has been developed. The discussions in Chapters II. and III. of infinite aggregates, and especially of the notion of the *power* or *cardinal number* of such an aggregate, will throw light upon this subject.

CHAPTER II.

THEORY OF SETS OF POINTS.

45. AN aggregate of real numbers, each element of which consists of a single real number, is defined by any prescribed set of rules or specifications which are of such a nature that, when any real number whatever is arbitrarily assigned, they theoretically suffice to determine whether such real number does or does not belong to the aggregate. The difficulty of regarding an aggregate, so defined, as a definite object, is bound up with the difficulties connected with the notion of the linear continuum, *i.e.* the aggregate of all real numbers, out of which the defined aggregate is to be obtained by a process of selection which, except in the case of a finite aggregate, can never be actually carried out in its entirety, but which is determined by a rule or set of rules. The precise scope of the definition will be rendered clearer by the consideration of various classes of actually defined aggregates which will be considered in the present Chapter; moreover, the theoretical difficulties of the notion of such an aggregate, in general, will be in some measure elucidated by the discussions in the present and the following Chapters, of the notion of the power, or cardinal number, of an aggregate.

In accordance with the principle explained in § 43, each number of a given aggregate may be represented by a single point on a fixed straight line; thus, to an aggregate of numbers, there corresponds an aggregate of points on the straight line. An aggregate of single numbers, or of their equivalent points, we shall speak of as a linear set of points.

The theory of linear sets of points, of which the present Chapter contains an account, arose historically from the discussion of questions connected with the theory of Fourier's series and of the functions which can be represented by such series. A consideration of the properties and peculiarities of the sets of points at which infinities or other discontinuities of such functions exist, led to a study of the properties of linear sets in general, and to the development by G. Cantor, P. Du Bois Reymond, Bendixson, Harnack, and others, of a general theory which has lately received wide applications both in Analysis and in Geometry.

Corresponding to the theory of linear sets of points, there exist theories of plane, solid, or n-dimensional sets of points. A set of points in n

dimensions is an aggregate each element of which is specified by n real numbers. The theory of such sets proceeds on lines similar to that of linear sets; indeed most of the investigations in the latter theory are capable of extension, with slight modification, to the more general cases. For the sake of brevity, we shall in general confine our attention to linear sets; some indications will, however, be given of the mode in which the definitions and properties which arise in the theory of linear sets, may be extended to the case of plane or solid sets; and a few properties peculiar to non-linear sets will be given.

The whole theory is fundamentally arithmetical; the geometrical representation and nomenclature being a matter of convenience, not of necessity.

THE UPPER AND LOWER BOUNDARIES OF A SET OF POINTS.

46. Let a set of points be such that every point of the set lies upon a straight line, the position of each point being determined by its distance from a fixed origin upon the straight line, in the manner explained in § 43. If a point β exists, such that no number of the set is greater than β, the set is said to be bounded on the right. In this case it will be shewn that there is a definite point b, such that no point of the set is on the right of b, and such that either b is itself a point of the set, or else points of the set are within the interval $(b-\epsilon, b)$ however small the positive number ϵ may be taken to be.

When b is a point of the set, there may or may not be other points of the set in every interval $(b-\epsilon, b)$. This point b is said to be the *upper limit* of the set, when points of the set lie within every interval $(b-\epsilon, b)$. In any case, when b is a point of the set, it is said to be the *upper extreme* point of the set. The term *upper boundary* may be applied to the point b, whether it be the upper limit or only the upper extreme point.

In case b is both the upper limit, and the upper extreme point, of the set, the upper limit is said to be attained; and b is then called the maximum point of the set.

To prove the existence*, under the condition stated, of an upper boundary, as above defined, it may be observed that all the numbers of the continuum of real numbers can be divided into two classes, one of which contains every number which is greater than all the numbers of the set, and the other of which contains every number which either belongs to the set or is less than some or all of the numbers of the set. The section thus specified defines a number b which is the upper boundary of the set.

* The existence of upper and lower boundaries was proved by Weierstrass, in his lectures. See also Bolzano, *Abh. d. Böhmischen Gesellsch. d. Wiss.*, vol. v, Prag, 1817.

In a similar manner, it may be shewn that, if the set is bounded on the left, *i.e.* if a point can be found such that all the points of the set are on the right of such point, then a point a exists, which is such that no points of the set are on the left of a, and such that either a is a point of the set, or else points of the set are within every interval $(a, a+\epsilon)$, where ϵ is an arbitrary positive number. Both conditions may be satisfied simultaneously.

In case points of the set lie within every interval $(a, a+\epsilon)$, then a is called the *lower limit* of the set; and the lower limit is said to be attained if a be itself a point of the set. In any case in which a is a point of the set, it is then said to be the *lower extreme* point of the set. The term *lower boundary* may in all cases be applied to a.

A set of points which has both an upper and a lower boundary is said to be a bounded set.

47. If no point b exists, which is either the upper limit, or the upper extreme point, of the set, then the set is said to be unbounded on the right; or it is said that the upper limit of the set is $+\infty$; the two statements being regarded as tautological. Similarly, if no lower limit nor lower extreme point a exists, the set is said to be unbounded on the left; or it is said that the lower limit is $-\infty$.

The symbols $+\infty, -\infty$ do not really represent numbers; they must be taken to represent what is sometimes spoken of as the improperly infinite, *i.e.* the mere absence of an upper or a lower boundary respectively. In order, however, to avoid circumlocution in the statement of theorems concerning sets, it is usually convenient to speak of $+\infty, -\infty$, used in the above sense, as if they were numbers which correspond to upper and lower limits respectively.

In the present Chapter, it will in general be assumed that the sets treated of are bounded; and the interval (a, b) will be said to be the interval in which the set exists. This restriction is not so great a one as might at first sight appear; for an unbounded set can be placed into correspondence with a bounded one, in such a manner that the relative order of any two points in the one set is the same as that of the corresponding points in the other set. If $x' = \dfrac{x}{\sqrt{x^2+1}}$, where the radical is taken to have always the positive sign, then to a point x, in the unlimited line $(-\infty, +\infty)$, there corresponds a point x', in the limited line $(-1, +1)$; and also $x_1' \gtreqless x_2'$, according as $x_1 \gtreqless x_2$. The same object might have been attained by using the transformation

$$x' = \frac{2}{\pi}\tan^{-1}x.$$

There is no real loss of generality in considering only such sets as lie in a given interval, say (0, 1); for the relation $x' = \frac{x-\alpha}{\beta-\alpha}$ establishes a complete correspondence between sets in the interval (α, β) and sets in the interval (0, 1), the relative order of points being preserved in the correspondence.

The points of the interval (α, β) may be made to correspond in order with the points of the interval (0, 1), in such a manner that an arbitrarily chosen point γ within (α, β), corresponds to an arbitrarily chosen point within (0, 1); for example the point $\frac{1}{2}$. This correspondence can be effected by the transformation

$$\frac{x'}{x'-1} = -\frac{x-\alpha}{x-\beta}\cdot\frac{\gamma-\beta}{\gamma-\alpha}.$$

LIMITING POINT OF A SET OF INTERVALS.

48. Let $(a_1, b_1), (a_2, b_2), \dots (a_n, b_n) \dots$ be an unending sequence of intervals which are such that any one of them (a_n, b_n) lies entirely in the preceding one (a_{n-1}, b_{n-1}), the two having at most one end-point common; thus $a_n \geqq a_{n-1}$, $b_n \leqq b_{n-1}$; moreover, suppose that the lengths $b_1 - a_1$, $b_2 - a_2, \dots$ $b_n - a_n, \dots$ form a sequence which converges to zero, the condition for which is that, corresponding to any arbitrarily small ϵ, n can be so chosen that $b_m - a_m$, for all values of m which are $\geqq n$, is $< \epsilon$. It will be seen that, in accordance with the axioms explained in § 43, *there exists one point and one only which is in every interval of the sequence.* This point may be called the *limiting point* of the sequence of intervals.

Each of the aggregates $(a_1, a_2, a_3, \dots a_n, \dots)$, $(b_1, b_2, \dots b_n, \dots)$ being convergent, defines a number; and in fact, in virtue of the definition of equality in § 25, they define the same number x. This number x is not less than a_n and not greater than b_n, whatever n may be; the point x therefore lies in all the intervals, and is the limiting point whose existence was to be remarked. If y be any number greater (or less) than x, we can find n so great that $b_n - x < y - x$, if $y > x$; or that $x - a_n < x - y$, if $x > y$: thus y does not lie in (a_n, b_n). Hence there is only one point which satisfies the prescribed conditions.

If for every n, from and after some fixed value, the inequalities $a_n > a_{n-1}$, $b_n < b_{n-1}$ both hold, then the limiting point x is in the interior of all the intervals of the sequence. If, from and after some fixed value of n, say n_1, we have $a_n = a_{n-1}$, $b_n < b_{n-1}$, the limiting point x coincides with the common end-points $a_{n_1-1}, a_{n_1}, a_{n_1+1}, \dots$.

THE LIMITING POINTS AND THE DERIVATIVES OF A LINEAR SET.

49. If a point x be taken in the interval (a, b), an interval $(x-\epsilon_1, x+\epsilon_2)$ which lies entirely in (a, b) is called a *neighbourhood* of the point x; and this neighbourhood may be made as small as we please by proper choice of ϵ_1 and ϵ_2. An interval $(x, x+\epsilon_2)$ is called a neighbourhood of x on the right, and $(x-\epsilon_1, x)$ is called a neighbourhood of x on the left. The end-points a and b can only have neighbourhoods on the right and the left respectively.

If a linear set of points not finite in number (denoted by G) is in the interval (a, b), then a point P, in whose arbitrarily small neighbourhood there exists at least one point of G not identical with P, is called a limiting point of the set G, whether P belongs to G or not.

The fundamental theorem will now be proved that *every set of an infinite number of points G, in an interval (a, b), possesses at least one limiting point.*

Divide (a, b) into m equal parts; then in one at least of these, say (a_1, b_1), there is an infinite number of points of G; and if this is the case in more than one of the parts, we may take any one of these for (a_1, b_1). Divide (a_1, b_1) into m equal parts; then there must be one of these parts at least, say (a_2, b_2), which contains an infinite number of points of G. Proceeding in this manner, we obtain a sequence of intervals (a_1, b_1), $(a_2, b_2) \ldots (a_n, b_n) \ldots$, of lengths $\frac{1}{m}(b-a)$, $\frac{1}{m^2}(b-a), \ldots \frac{1}{m^n}(b-a), \ldots$ in each of which there is an infinite number of points of G. In accordance with the theorem of § 46, there exists one point x which is in all the intervals (a_n, b_n). Take any arbitrarily small neighbourhood of x, say $(x-\epsilon_1, x+\epsilon_2)$; then if n be chosen so large that $\frac{1}{m^n}(b-a)$ is less than the smaller of the numbers ϵ_1, ϵ_2, the interval (a_n, b_n) lies entirely within $(x-\epsilon_1, x+\epsilon_2)$. Hence in the arbitrarily small neighbourhood of x, there is an infinite number of points of G; therefore x is a limiting point of G.

It has thus been shewn that G has at least one limiting point. It may have a finite number, or an indefinitely great number, of limiting points. It should be observed that a limiting point of G may or may not itself be a point of G. If either boundary of the set be not a point of G, then it is certainly a limiting point of G; it may however be both.

A limiting point P of a set G is a limiting point on both sides, if an indefinitely great number of points of G lie in every neighbourhood of P on the right, and also in every neighbourhood of P on the left. Otherwise P is a limiting point of G on one side only.

50. In the case of a set for which either the upper boundary or the lower boundary is absent, or both are absent, we may use either method given in § 47, of making the set correspond with a bounded set in the interval $(-1, +1)$. To any definite interval in $(-\infty, +\infty)$, there corresponds a definite interval in $(-1, +1)$, neither end-point of which is at -1 or 1. To a limiting point interior to $(-1, +1)$, there corresponds a limiting point in $(-\infty, +\infty)$. For if x' be such a limiting point of the set in $(-1, +1)$, there are in any neighbourhood of x' an infinite number of points of the set; and to this neighbourhood there corresponds a neighbourhood of the corresponding point x of the unbounded set. Thus the point x is such that, in any neighbourhood of it, there are an infinite number of points of the set, so that x is a limiting point. The only case in which the unbounded set has no limiting point, is when the corresponding bounded set has for its sole limiting points the end-points $-1, +1$, or one only of these; and in this case we may say that ∞, $-\infty$, or one of these, is the *improper* limiting point of the unbounded set. The properties of an unbounded set in relation to its limiting points are thus not essentially different from those of a bounded set.

51. Returning to the case of a set G in an interval (a, b), we observe that the limiting points of G form a set of points which may be finite or infinite; this set is called the *derived set**, or *first derivative* of G, and may be denoted by G'. In case the set G' contains an infinite number of points, it possesses itself a derivative set G'', which is called the *second derivative* of G. If we proceed in this manner, we may obtain a series

$$G', G'', G''', \ldots G^{(n)}$$

of derivatives of G. If the nth derivative $G^{(n)}$ contains a finite number only of points, then these have no limiting point, and we may say that $G^{(n+1)} = 0$. It may however happen that, however large the integer n may be, the derivative $G^{(n)}$ contains an indefinitely great number of points; and thus a next derivative exists.

A set G which possesses only a finite number of derivatives is said to be of the first species.

In this case, if $G^{(s)}$ contains only a finite number of points, the set G is said to be of order† s. Thus, for example, a set of the first species and order zero, contains only a finite number of points; and a set of the first species and order 1, has a first derivative which contains only a finite number of

* The notion of the derivative of a set was introduced by Cantor, *Math. Annalen*, vol. v (1872), p. 128. Du Bois Reymond contemplated the existence of limiting points of various orders, *Crelle's Journal*, vol. LXXIX (1874), p. 30; in *Math. Annalen*, vol. XVI, p. 128, Du Bois Reymond defined a limiting point of infinite order.

† Cantor, *Math. Annalen*, vol. v, p. 129.

points. It will be observed that the order of each derivative of G is less by unity than that of the one which precedes it.

A set G which possesses an indefinite number of derivatives is said to be of the second species.

As an example, we may consider the set of rational numbers in the interval $(0, 1)$. The first derivative of this set contains every real number in $(0, 1)$, and all subsequent derivatives are identical with the first.

The theorem, that every non-finite linear set of points possesses a limiting point, is a particular case of the theorem that every non-finite set in a finite portion of an n-dimensional continuum has a limiting point. In this case we may take the neighbourhood of a point to be either a "sphere" of arbitrarily small radius ρ, or a "rectangular cell" with sides parallel to the coordinate axes, and the point at the centre. The space in which the set of points exists can be divided into a finite number of overlapping "spherical" or of "square" portions, and the argument then proceeds as in the case of a linear set.

EXAMPLES.

1. Let*

$$G = \left(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4} \cdots \frac{1}{n} \cdots\right).$$

We see that G' consists of the single point O, which does not belong to G; thus G is of the first species and of order 1.

2. Let† the points of G be given by

$$\frac{1}{3^{s_1}} + \frac{1}{5^{s_2}} + \frac{1}{7^{s_3}} + \frac{1}{11^{s_4}},$$

where s_1, s_2, s_3, s_4 each have all positive integral values. Here G' consists of the four sets of points given by

$$\frac{1}{3^{s_1}} + \frac{1}{5^{s_2}} + \frac{1}{7^{s_3}}, \quad \frac{1}{3^{s_1}} + \frac{1}{5^{s_2}} + \frac{1}{11^{s_4}}, \quad \frac{1}{3^{s_1}} + \frac{1}{7^{s_3}} + \frac{1}{11^{s_4}}, \quad \frac{1}{5^{s_2}} + \frac{1}{7^{s_3}} + \frac{1}{11^{s_4}};$$

and of the six sets of points

$$\frac{1}{3^{s_1}} + \frac{1}{5^{s_2}}, \quad \frac{1}{3^{s_1}} + \frac{1}{7^{s_3}}, \quad \frac{1}{3^{s_1}} + \frac{1}{11^{s_4}}, \quad \frac{1}{5^{s_2}} + \frac{1}{7^{s_3}}, \quad \frac{1}{5^{s_2}} + \frac{1}{11^{s_4}}, \quad \frac{1}{7^{s_3}} + \frac{1}{11^{s_4}};$$

and of the four sets of points

$$\frac{1}{3^{s_1}}, \quad \frac{1}{5^{s_2}}, \quad \frac{1}{7^{s_3}}, \quad \frac{1}{11^{s_4}};$$

together with the single point O. G'' consists of the last ten of these sets, and of the point O. The second derivative G''' consists of the last four sets, and of the point O; G'''' consists of the point O only. The set G is of the first species and of the fourth order.

* Cantor, *Math. Annalen*, vol. v (1872).

† Ascoli, *Ann. di Mat.*, Series II, vol. VI, p. 56, 1875.

3. Let* the points of G be given by

$$\frac{1}{a_1}+\frac{1}{a_2}+\dots+\frac{1}{a_n},$$

where n is a fixed number, and each of the numbers $a_1, a_2, \dots a_n$ takes every positive integral value. In this case G is of order n.

4. The zeros† of the function $\sin\frac{1}{x}$ form a set similar to that in Example 1. The zeros of the function $\sin\left(\frac{1}{\sin\frac{1}{x}}\right)$ form a set of the second order, those of $\sin\left(\frac{1}{\sin\frac{1}{\sin\frac{1}{x}}}\right)$ form a set of the third order, and so on.

5. Let‡ the points of G be given by

$$\frac{1}{2^{m_1}}+\frac{1}{2^{m_1+m_2}}+\dots+\frac{1}{2^{m_1+m_2+\dots+m_n}},$$

where $m_1, m_2, \dots m_n$ have all positive integral values, including zero, and n is a fixed integer. It can be seen that $G^{(n)}$ consists of the point zero only.

THE DISTRIBUTION OF POINTS OF A SET IN THE INTERVAL.

52. If $G_1, G_2, G_3 \dots G_n$ denote a number of sets of points, a set which contains all points which belong to any one or more of the given sets is called their *common measure*, and is denoted by $M(G_1, G_2, \dots G_n)$. That set which contains all those points which belong to every one of the given sets is called their *greatest common divisor*, and may be denoted by $D(G_1, G_2, \dots G_n)$.

A set of points is said to be an isolated set, when no point of the set is a limiting point. Thus if G be such a set, we have $D(G, G')=0$.

If from any set we remove those points which also belong to its derivative, the remainder forms an isolated set; thus $G-D(G, G')$ forms an isolated set. Any set may be regarded as the sum of an isolated set and of a set which is a divisor of the derivative.

A set, all of whose limiting points belong to the set itself, is said to be closed. Thus in a closed set G, the derivative G' is a divisor of G.

A set which is such that every point of the set is a limiting point is said to be dense-in-itself.

For a set G, dense-in-itself, G is a divisor of the derivative G'. The rational numbers in (0, 1), form an example of a set which is dense-in-itself.

* H. J. S. Smith, *Proc. Lond. Math. Soc.*, vol. VI, p. 145, 1875.

† P. Du Bois Reymond, *Journ. f. Math.*, vol. LXXIX, p. 36.

‡ Mittag-Leffler, *Acta Math.*, vol. IV, p. 58.

*A set G which is both closed and dense-in-itself is said to be perfect**. Thus a perfect set G is identical with its derivative. It follows that every perfect set is of the second species.

By some writers† the term *perfect* is applied to sets which, in accordance with the terminology of Cantor here adopted, are only closed, without necessarily being dense in themselves; what we call a perfect set is then spoken of as an absolutely‡ perfect set.

If in an interval (a, b), a smaller one (a', b') such that $a \leqq a'$, $b \geqq b'$ is taken, then the latter may be called a sub-interval of the former interval.

If in the interval (a, b), in which a set of points G is contained, no sub-interval whatever, however small, can be found which does not contain points of G, then the set G is said to be everywhere-dense§, or simply, dense in the interval (a, b).

By Du Bois Reymond‖, the term *pantachisch* was used with the same meaning as *everywhere-dense.*

A set which is everywhere-dense is also dense-in-itself, but the converse does not hold.

It will be seen that, if G is everywhere-dense in (a, b), every sub-interval of (a, b) must contain an indefinitely great number of points of G. The derivative G' of G must contain every point in the interval (a, b), since the arbitrarily small neighbourhood of any point whatever of (a, b) contains an indefinitely great number of points of G; and therefore every point must be a limiting point. This property, that G' contains every point of (a, b), may be used as the definition¶ of an everywhere-dense set.

If, in every sub-interval (a', b') of the interval in which G exists, a part (a'', b'') can be found which contains no points of G, then G is said to be nowhere-dense, or non-dense in (a, b).

An example of a set which is everywhere-dense in its interval is the set of rational numbers in the interval $(0, 1)$. A set which is everywhere-dense in its interval, or in any sub-interval, is necessarily of the second species; a set of the first species is nowhere-dense in its interval.

A set which is in no sub-interval dense-in-itself, is said to be separated.

If in an interval (a, b), an indefinitely great number of sub-intervals, which may or may not overlap, be taken, and no sub-interval (α, β) of (a, b) can be found which is wholly external to all the given sub-intervals, then the set of sub-intervals is said to be everywhere-dense in (a, b).

* Cantor, *Math. Annalen*, vol. XXI.

† For example Jordan, see *Cours d'Analyse*, vol. I, p. 19.

‡ Borel, *Leçons sur la théorie des fonctions*, p. 36.

§ Cantor, *Math. Annalen*, vol. XV, p. 2.

‖ *Math. Annalen*, vol. XV, p. 287.

¶ Baire, *Annali d. Mat.*, Series 3, vol. III, p. 29.

53. The following fundamental theorem will now be proved:

All the derivatives G', G'', G''', ... $G^{(n)}$, ... of a given set G are closed sets, and each of these derivatives, after the first, consists only of points belonging to the preceding one, and therefore to G'.

If a point P of $G^{(n)}$, where $n \geqq 2$, existed, which did not belong to G', then a neighbourhood of P could be found, so small as to contain only a finite number of points of G, or no such points; and this neighbourhood would therefore contain no points of G', and therefore none of G'', G''', ... $G^{(n)}$; which would be contrary to the hypothesis that P belongs to $G^{(n)}$. Therefore every point of $G^{(n)}$ ($n \geqq 2$) belongs to G'; and consequently G' is a closed set. If we take $G^{(n-2)}$ to be the original set, it follows from the above that every point of $G^{(n)}$, the second derivative, belongs to $G^{(n-1)}$ the first derivative. We have thus shewn that $G^{(n)}$ is the greatest common divisor of G', G'', ... $G^{(n)}$; that is

$$G^{(n)} = D(G', \ G'', \dots G^{(n)}).$$

The derivative G', of a set G which is dense-in-itself, is perfect.

For G' is closed, and every point of G belongs to G'; thus G' contains no point which is not a limiting point of the set G'. Therefore G' is dense-in-itself; and hence it is perfect.

A set G_1, which consists of some, but not all, of the points of G, is said to be a component of G.

If the component G_1 of G, be such that every point of G is a limiting point of G_1, then the component G_1 is said to be everywhere-dense relatively to, or simply, dense in G.

In the case in which G is the continuum (a, b), this definition agrees with that which has been given for a set which is everywhere-dense in the interval in which it is contained.

ENUMERABLE AGGREGATES.

54. *An aggregate which contains an indefinitely great number of elements is said to be enumerable*, or countable* (abzälbar, dénombrable), *when the aggregate is such that a* (1, 1) *correspondence can be established between the elements and the set of integral numbers* 1, 2, 3, ...

An aggregate of objects is therefore enumerable if the objects can be arranged in a series which has a first term, and in which any assigned object belonging to the aggregate has a definite place assigned by a definite ordinal number n. Thus the elements of an enumerable aggregate can be represented by a series of terms

$$u_1, \ u_2, \ \dots \ u_n, \ \dots.$$

* Cantor, *Crelle's Journal*, vol. LXXVII (1873), p. 258.

It follows from this definition that the elements of two enumerable aggregates are such that a (1, 1) correspondence can be established between them.

If a new aggregate be formed by selecting elements from those which belong to an enumerable aggregate, an indefinitely great number of such elements being taken, then the new aggregate is also enumerable. For such an aggregate selected from $u_1, u_2, \dots u_n, \dots$ is $u_r, u_s, u_t \dots (r < s < t \dots)$, which satisfies the conditions of having a first term, and of having each element of the aggregate in a definite place in the series. It thus appears, that a (1, 1) correspondence can be established between an enumerable aggregate and one which is a part of that aggregate, provided this part be not finite. This is the characteristic property which distinguishes an aggregate containing an indefinitely great number of elements from one containing only a finite number of elements. For example, a (1, 1) correspondence exists between all the integral numbers and all the odd numbers, or between all the integral numbers and all the prime numbers.

*If a finite number of enumerable aggregates be taken, or even if the number of such aggregates be indefinitely great, but enumerable, then the new aggregate formed by the whole is itself enumerable**.

We may denote such a composite aggregate by the letters

$$\begin{array}{l} u_{11},\ u_{12},\ u_{13},\ \dots\ u_{1n},\ \dots \\ u_{21},\ u_{22},\ u_{23},\ \dots\ u_{2n},\ \dots \\ u_{31},\ u_{32},\ \dots\dots\dots\dots \\ \dots\dots\dots\dots\dots\dots \\ \dots\dots\dots\dots\dots\dots \end{array}$$

and we shall shew that the double series so formed represents an enumerable aggregate. To see this, it is sufficient to write the series in the form

$$\begin{array}{llll} u_{11} & & & \\ u_{12}, & u_{21} & & \\ u_{13}, & u_{22}, & u_{31} & \\ \dots\dots\dots & & & \\ u_{1,n-1}, & u_{2,n-2}, & u_{3,n-3}, & \dots\ u_{n-1,1} \\ \dots\dots\dots\dots & & & \end{array}$$

where the sum of the indices is the same for all the terms which are written in one horizontal line. It is now clear that each number u_{pq} has a definite place in a series in which u_{11} has the first place; the double series is therefore enumerable.

* Cantor, *Crelle's Journal*, vol. LXXXIV (1875).

An important particular case of the above theorem is the following theorem:

The aggregate of all the rational numbers is enumerable.

A rational number p/q may be denoted by $u_{p,q}$: therefore the aggregate is enumerable. It makes no difference that any particular number p/q occurs an indefinite number of times as $u_{rp,rq}$; since, if all such terms except those for which $r = 1$, and p/q is in its lowest terms, be removed, the aggregate left is still enumerable.

55. A more general theorem has also been established by Cantor*. An algebraical number is one which is a root of an algebraical equation in which the coefficients are all rational numbers, so that the coefficients may without loss of generality be taken to be integers. Cantor's theorem is, that *all the algebraical numbers form an enumerable aggregate.*

To prove this theorem, let

$$p_0x^n + p_1x^{n-1} + \ldots + p_n = 0$$

be an equation in which $p_0, p_1 \ldots p_n$ are all positive or negative integers; and let

$$|p_0| + |p_1| + \ldots + |p_n| + n = N;$$

then N is a positive integer which may be called the rank of the equation. It is clear that there are only a finite number of equations of any given rank, these equations having only a finite number of roots. If then we let $N = 3, 4, 5, \ldots$ successively, we can arrange all the algebraical numbers in a simple series; and thus they form an enumerable aggregate. The aggregate which is formed of all the real algebraical numbers is also itself enumerable.

A number which is not an algebraical number is said to be transcendental. The existence of transcendental numbers was first established† by Liouville, who shewed how examples of such numbers could be formed. No general criterion is known by which it can be decided whether a number, defined by a given analytical procedure, is algebraical or transcendental. The first case in which such a number, well known in Analysis, was shewn to be transcendental was that of the number e, the base of the natural system of logarithms; and the first proof that e is transcendental was given by Hermite. The next case in which a number defined analytically was shewn to be transcendental was that of the number π. The first demonstration of this important fact is due to Lindemann‡, who proved the more general theorem that, if $e^x = y$, the two numbers x, y cannot both be algebraical, except in the case $x = 0$, $y = 1$. It follows that the natural logarithms of all algebraical

* *Crelle's Journal*, vol. LXXVII.

† *Liouville's Journal*, vol. XVI, 1851.

‡ See *Math. Annalen*, vol. XX.

numbers are transcendental, as also all numbers of which the natural logarithms are algebraical.

56. The following fundamental theorem will now be established* :—

The aggregate which consists of the continuum of numbers in a given interval is not enumerable.

Suppose $\omega_1, \omega_2, \omega_3, \ldots$ denote the numbers in an enumerable aggregate; it will then be shewn that, between any two numbers α, β as near as we please, a number occurs which does not belong to the enumerable aggregate. It will then follow that in the given interval there is an unlimited number of points which do not belong to the enumerable aggregate, and thus that the latter cannot contain all the points of the continuum. If the enumerable set of points is not everywhere-dense in (α, β), then smaller sub-intervals inside (α, β) can be taken which contain no points of the aggregate; and thus we have only to consider the case in which the given aggregate is everywhere-dense in (α, β). Let ω_{κ_1} be the first of the points $\omega_1, \omega_2, \ldots$ which lies within (α, β), and ω_{κ_2} be the next of these points which lies within (α, β), so that $\kappa_1 < \kappa_2$. Let α' be the smaller, and β' the greater of the numbers $\omega_{\kappa_1}, \omega_{\kappa_2}$, then $\alpha < \alpha' < \beta' < \beta$, and $\kappa_1 < \kappa_2$; and if $\mu < \kappa_2$, then ω_μ does not lie within the interval (α', β'). Considering this latter interval, let $\omega_{\kappa_3}, \omega_{\kappa_4}$ be the first two of the numbers of the enumerable aggregate which lie within (α', β'), and let α'' be the smaller and β'' the greater of these; then $\alpha' < \alpha'' < \beta'' < \beta'$, and $\kappa_2 < \kappa_3 < \kappa_4$. Proceeding in this manner we obtain a whole series of sub-intervals each one of which is entirely within the preceding one; thus $(\alpha^{(\nu)}, \beta^{(\nu)})$ lies within $(\alpha^{(\nu-1)}, \beta^{(\nu-1)})$; and if $\mu \gtreqless \kappa_{2\nu}$, then ω_μ does not lie within $(\alpha^{(\nu)}, \beta^{(\nu)})$; also

$$\kappa_1 < \kappa_2 < \kappa_3 \ldots < \kappa_{2\nu-2} < \kappa_{2\nu-1} < \kappa_{2\nu},$$

and $2\nu \leqq \kappa_{2\nu}$; and thus ω_ν lies outside $(\alpha^{(\nu)}, \beta^{(\nu)})$. Since the numbers $\alpha', \alpha'', \alpha''' \ldots$ are in ascending order, and all lie within (α, β), they have a limit A; similarly $\beta', \beta'', \beta''' \ldots$ have a limit B; and $\alpha^{(\nu)} < A \leqq B < \beta^{(\nu)}$. If $A < B$, then since all the numbers ω_ν are outside the interval (A, B), the given aggregate is not everywhere-dense in (α, β); which is contrary to hypothesis. Hence we have $A = B$, and the number A or B, is a number which does not occur in the aggregate $\omega_1, \omega_2, \ldots$; which was what we had to prove.

It will be observed that the point of the foregoing proof consists in the fact, that an everywhere-dense enumerable aggregate necessarily has limiting points which do not belong to the aggregate.

A second proof† that the continuum is not enumerable is the following :— Without loss of generality, the interval may be taken to be (0, 1). Suppose it to be possible to arrange all the numbers in this interval in order, so that there is a first, a second, a third and so on; and so that every number occurs

* Cantor, *Crelle's Journal*, vol. LXXVII.

† *Jahresbericht der deutschen Math. Vereinig.* vol. I, p. 77.

somewhere in the arrangement. Let the numbers, in order, be exhibited as decimals

$$\cdot p_{11}\, p_{12}\, p_{13} \cdots\cdots$$
$$\cdot p_{21}\, p_{22}\, p_{23} \cdots\cdots$$
$$\cdot p_{31}\, p_{32}\, p_{33} \cdots\cdots$$
$$\cdots\cdots\cdots\cdots$$
$$\cdots\cdots\cdots\cdots$$

where each p stands for a digit 0, 1, 2, ... 9, and numbers, in which the digits, from and after some fixed place, are all 9, are excluded; then if a number can be defined which does not occur in the above series, a contradiction will have been shewn to be involved in the supposition that all the numbers can be exhibited in the above manner. Now this can be done; for the number $\cdot(p_{11})(p_{22})(p_{33})\ldots$ where (p) denotes any digit except p, say $p+1$ or 0, according as $p < 9$, or $p = 9$, differs in at least one place of the decimal, from every number in the above set; and the contradiction is thus established.

THE POWER OF AN AGGREGATE.

57. A notion of fundamental importance in the theory of aggregates is that of the *power* of an aggregate. This notion will be considered more generally and fully in the next Chapter, where it will be shewn that the power of an aggregate is the generalization of the notion contained in the cardinal number of a finite aggregate. At present, an account of the notion of the power of an aggregate will be given, so far as it is necessary for the application to the case of sets of points.

Two aggregates of objects are said to have the same power, or cardinal number, when a (1, 1) correspondence can be established between them, so that each element of either of the aggregates corresponds to one single element of the other.

Finite aggregates have the same power when they consist of the same number of elements, *i.e.* when they have the same cardinal number. Of aggregates which are not finite we consider first enumerable aggregates. Every enumerable aggregate has the power of the aggregate of integral numbers; and this we may denote by a. It has been shewn above that, if from an aggregate of power a any elements be removed, then the remaining aggregate, provided it contains a non-finite number of elements, has still the same power a. It has further been shewn that the composite aggregate formed of a finite, or enumerable, number of enumerable aggregates has the same power a. It follows, as an interesting case, that the set of all those points of an n-dimensional space whose coordinates are rational numbers has the power a of the set of integral numbers, or of the rational numbers in a given linear interval.

It is easily shewn that the power of the set of all the points in an interval (a, b) is the same as that in any other finite interval, say $(0, 1)$; for $\frac{x-a}{b-a} = x'$ establishes a $(1, 1)$ correspondence between the points x of (a, b) and the points x' of $(0, 1)$. Again the relation $\frac{x}{\sqrt{x^2+h^2}} = x'$, establishes a $(1, 1)$ correspondence between all real numbers, and those in the interval $(-1, 1)$; and thus the power of all real numbers is the same as that of all those in a finite interval. This power is called the power of the continuum, and may be denoted by c.

As regards unenumerable aggregates in general, it can be shewn that the power of such an aggregate is unaltered by removing from the aggregate any elements which form an enumerable aggregate. Let A denote the given aggregate, and α the enumerable aggregate which is removed; and let B denote the remaining aggregate, which cannot be enumerable, for otherwise (α, B), or A, would be so also. From B, suppose an enumerable aggregate α' to be removed, leaving the aggregate C, thus $A = (\alpha, \alpha', C)$, $B = (\alpha', C)$. Now (α, α') and α', being both enumerable, have the same power; and a $(1, 1)$ correspondence therefore exists between their elements; and since A and B have the aggregate C in common, it therefore follows that A and B have the same power. As an example of this theorem, we see that the set of irrational points in a given interval has the power c of the set of all numbers in the interval. Again the set of transcendental numbers in a given interval has the power c of the continuum; whereas the set of algebraical numbers in the same interval has the power a.

The known infinite sets of points defined in accordance with the methods usual in the theory of sets of points, in a line or in a continuum of any number of dimensions, have either the power a or the power c; but it has not yet been established that every possible set of points has one of these two powers. Other aggregates have been contemplated which have a power higher than c; these will be referred to later, in dealing with the theory of functions.

58. *The n-dimensional continuum has the power c of the one-dimensional continuum*.*

To prove this theorem, we use the fact that any irrational proper fraction can be exhibited as an infinite continued fraction

$$x = \frac{1}{\alpha_1 +} \frac{1}{\alpha_2 +} \dots \frac{1}{+ \alpha_n +} \dots,$$

* Cantor, *Crelle's Journal*, vol. LXXXIV.

where $\alpha_1, \alpha_2, \alpha_n, \ldots$ are determinate integers for any given value of x. Let

$$x_1 = \frac{1}{\alpha_1+}\,\frac{1}{\alpha_{n+1}+}\,\frac{1}{\alpha_{2n+1}+}\ldots,$$

$$x_2 = \frac{1}{\alpha_2+}\,\frac{1}{\alpha_{n+2}+}\,\frac{1}{\alpha_{2n+2}+}\ldots,$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$x_n = \frac{1}{\alpha_n+}\,\frac{1}{\alpha_{2n}+}\,\frac{1}{\alpha_{3n}+}\ldots,$$

thus, corresponding to any value of x, a set of irrational numbers $x_1, x_2, \ldots x_n$ is uniquely determined, and conversely to any set of irrational numbers $x_1, x_2, \ldots x_n$, a value of x is uniquely determined.

It has thus been shewn that the irrational points of the linear continuum (0, 1) correspond uniquely to those points of the n-dimensional continuum in which each coordinate is in the interval (0, 1), and is irrational. It has been shewn in § 57, that the set of irrational values of x_1, in the interval (0, 1) has the same power as the set of all the numbers in this interval. Since this holds also for $x_2, x_3, \ldots x_n$, it follows that a (1, 1) correspondence can be established between that set of points in the n-dimensional continuum, for which $x_1, x_2, \ldots x_n$ all have irrational values, and the set in which $x_1, x_2, \ldots x_n$ have all values rational or irrational; thus these sets have the same power. Hence the set of all points of the n-dimensional continuum, in which each coordinate is in the interval (0, 1), has the same power as the set of all points in the linear interval (0, 1). It has thus been shewn that the n-dimensional continuum has the same power c as that of one dimension.

THE ARITHMETIC CONTINUUM.

59. The arithmetic continuum having been obtained by adjoining to the set of rational numbers the set of all their limiting points, the question arises how far it is legitimate to consider the complete set so obtained as constituting a single object determined by means of the elements of which it is composed. A finite set of numbers, or points, constitutes a single object determined by means of its parts, in the sense, that those parts can be exhaustively exhibited by means of a finite number of specifications representable by a finite number of symbols. An enumerable set of numbers, or of points, in particular the set of rational numbers, is not determinate in the sense that the elements of the set can be exhaustively exhibited; but it is determinate in the sense that a table can be formed in which each particular number of the set occupies a determinate place; and each particular number can be represented by means of a finite number of symbols. Such a set may be regarded as an aggregate, or single object, in the same sense in which the natural numbers 1, 2, 3, ... may be regarded as forming an aggregate. When we come, however, to the case of the continuum, or aggregate of all real numbers, the fact that this aggregate is unenumerable introduces a new element into the question of the legitimacy of considering the set of these numbers as forming a determinate whole, or as constituting a single object of thought. The set of real numbers cannot be tabulated in such a manner that no number fails to occur at some definite place in the table. No set of rules or specifications can be given which suffice to determine successively all the numbers of the set, and no finite set of symbols can exhaustively exhibit the numbers. The only sense in which the numbers of the set are determinate is that each such number is the limit of a convergent sequence of numbers, taken from the unending table formed by the rational numbers. It may fairly be doubted whether such a negative specification of elements amounts to a valid synthetical definition of a determinate aggregate; this point will however be further discussed in Chapter III., in connection with the general theory of aggregates. It will there be shewn that the arithmetic continuum has an order-type possessing definite characteristics which, in their totality, uniquely characterize it. This expresses the only kind of unity which can appertain to the continuum, considered as an arithmetic construction. If it be held that we possess an independent knowledge of the existence of the geometrical continuum, derived from our intuition of space, we may regard the function of the set of real numbers to consist, not in a synthetical formation of the concept of the continuum, but inversely in an analysis of the contents of the continuum. It is difficult to see how precision can be introduced into the intuitional notion of the continuum apart from some theory relating either to points or to infinitesimals; and the language employed in such a theory must be of a symbolical character amounting to the use of some kind of arithmetical notation. Regarding the geometrical continuum in this way as a single object of which we have a direct knowledge obtained from our intuitions of space and time, the reduction to a precise abstract form may be regarded as being made upon the assumption that the system of rational numbers, with their limits adjoined, is adequate to the analytical description of the continuum, in the sense that each point in the continuum is represented uniquely by a single real number, and that there is no point in the continuum which is not so represented. This amounts to a definition, in a certain sense, of the contents of the geometrical continuum. Such definition is not necessarily the only possible definition, but it is a legitimate one, provided it suffices for the purposes we have in view in Analysis and Geometry, and provided it does not conflict with the concept of Continuity as derived from intuition. The *generic* distinction between a continuous geometrical object, and a point, or set of points, situated in that object, is not capable of direct arithmetic representation. This does not, however, impair the efficiency of Arithmetical Analysis in dealing with geometrical objects. In Cartesian geometry, for example, Analysis is really concerned only with the points that can be determined in the geometrical objects with which it deals. This does not mean that a continuous geometrical object is analysed into points which are of necessity to be regarded as its "parts."

TRANSFINITE ORDINAL NUMBERS.

60. The theory of transfinite ordinal numbers had its origin* in the investigation of the theory of sets of points. The general abstract theory of such numbers, or order-types, will be deferred until the next Chapter; it is necessary however to introduce here the conceptions connected with the formation of these numbers, with a view to utilizing them in the theory of sets of points.

Let $P_1, P_2, \dots P_n, \dots$ denote a sequence of points in a given interval, representing a sequence $a_1, a_2, a_3, \dots$ of increasing numbers, so that

$$a_1 < a_2 < a_3 \dots < a_n \dots .$$

This sequence of points has a limiting point which is not one of the points of the sequence, and is on the right of all those points; this limiting point we may denote by P_ω. The symbol ω may be regarded as denoting a new ordinal number which comes after all the ordinal numbers $1, 2, 3, \dots n, \dots$; it is called the *first transfinite ordinal number*. The number ω is not contained in the sequence of finite ordinal numbers, but comes after all of them; and we shall see that it may be taken as the first of a new sequence of ordinal numbers, all of which must be regarded as ordinally greater than the finite ordinal numbers.

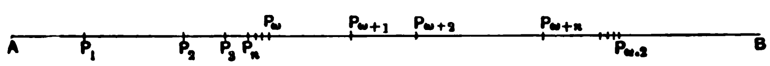


FIG. 1.

Suppose that beyond the point P_ω there are other points which we wish to regard as belonging to the same set as the points $P_1, P_2, \dots P_n, \dots P_\omega$; then these points will be denoted by $P_{\omega+1}, P_{\omega+2}, \dots P_{\omega+n}, \dots$; and if these points are finite in number, there will be one of them $P_{\omega+m}$ which is the last on the right. The indices of all the points of the set will be then

$$1, 2, 3, \dots n, \dots \omega, \omega+1, \omega+2, \dots \omega+m;$$

and the numbers $\omega, \omega+1, \dots \omega+m$ are regarded as a set of transfinite ordinal numbers, which commences with the first transfinite ordinal number ω, and contains the m succeeding transfinite ordinal numbers. It may however happen that the set of points $P_\omega, P_{\omega+1}, P_{\omega+2}, \dots$ has no last point. In that case, assuming that the points are all contained in a finite interval, the set

* An account of Cantor's earliest presentation of this subject will be found in *Math. Annalen*, vol. XXI.

has a limiting point which is not contained in the set itself; and this limiting point we denote by $P_{\omega+\omega}$ or $P_{\omega . 2}$, where $\omega . 2$ is an ordinal number which is not contained in the set $\omega, \omega + 1, \omega + 2, \ldots$, but comes after the numbers of the set.

If we wish to include further points which are on the right of $P_{\omega . 2}$, we must introduce numbers denoted by $\omega . 2 + 1, \omega . 2 + 2, \ldots$; and, in case these points form an infinite set in a finite interval, they will have a limiting point which will be denoted by $P_{\omega . 2+\omega}$ or $P_{\omega . 3}$. We have now the ordinal numbers

$$1, 2, 3, \ldots \omega, \omega + 1, \omega + 2, \ldots \omega . 2, \omega . 2 + 1, \omega . 2 + 2, \ldots \omega . 3.$$

If we proceed further in this manner it is clear that we may require numbers $\omega . n, \omega . n + 1, \omega . n + 2, \ldots \omega . \overline{n + 1}, \ldots$, where n denotes any finite number.

Further, it may happen that the set of points $P_{\omega}, P_{\omega . 2}, P_{\omega . 3}, \ldots P_{\omega . n}, \ldots$ is itself infinite, and has a limiting point on the right of all these points. This point we denote by P_{ω^2}; and the number ω^2 we consider to be a new ordinal number which succeeds all the numbers $\omega . n + m$, where n and m have all possible finite values.

Points on the right of P_{ω^2} may be denoted by means of the indices $\omega^2 + 1, \omega^2 + 2, \omega^2 + 3, \ldots$; and if these points are infinite in number, they may have a limiting point $P_{\omega^2+\omega}$.

Points on the right of $P_{\omega^2+\omega}$ may be denoted by the indices $\omega^2 + \omega + 1$, $\omega^2 + \omega + 2, \ldots$; if these have a limiting point, it will be denoted by the index $\omega^2 + \omega . 2$. Proceeding in this manner, we may have points of which the indices are $\omega^2 + \omega . 3, \omega^2 + \omega . 4, \ldots$. If there is an infinite set of such points, and the set has a limiting point, on the right of the set, this limiting point will have $\omega^2 + \omega^2$ or $\omega^2 . 2$ for its index.

If we proceed still further, we see as before that we may have to contemplate numbers of the form $\omega^2 . p + \omega . q + r$, where p, q, r are finite; afterwards $\omega^3, \omega^3 + 1, \ldots, \omega^3 . p + \omega^2 . q + \omega . r + s$, &c. The general type of ordinal numbers which can be obtained in this manner is represented by $\omega^n . p_n + \omega^{n-1} . p_{n-1} + \ldots + \omega . p_1 + p_0$; and it is clear that, for the representation of points of a given set, such numbers may be required as indices.

It may happen that the set of points whose indices are $\omega, \omega^2, \omega^3, \ldots$ is not finite; then the limiting point of such set will be denoted by the index ω^ω. Starting afresh with this number, we may form numbers such as

$$\omega^{\omega^n . p_n + \omega^{n-1} . p_{n-1} + \ldots + p_0}.$$

If the points whose indices are $\omega^\omega, \omega^{\omega^2}, \omega^{\omega^3}, \ldots$ do not form a finite set, their limiting point will be denoted by ω^{ω^ω}.

In a similar manner we may denote by ϵ_1 the number which comes after the sequence ω, ω^ω, ω^{ω^ω}, $\omega^{\omega^{\omega^\omega}}$ ...; and starting from ϵ_1, we may similarly proceed to form further numbers in endless succession.

61. All the ordinal numbers which can be formed in the manner above described are formed by means of the application of Cantor's two principles of generation (Erzeugungsprinzipien).

(1) *After any number another immediately succeeding it is formed by the addition of unity.*

(2) *After any endless sequence of numbers, a new number is formed which succeeds all the numbers in the sequence, and has no number immediately preceding it.*

All transfinite ordinal numbers which can be formed by means of these two principles of generation are said to be ordinal numbers of the *second class.* The finite ordinal numbers are said to be of the *first class*; they are formed successively, starting with the number 1, by means of the first principle of generation alone.

The numbers of the second class are of two essentially distinct species: (1) *non-limiting numbers,* those numbers which have each a number immediately preceding them, and from which they are formed by the addition of unity; for example $\omega+n$, $\omega^2.p+\omega.q+1$, $\omega^\omega+\omega+1$: and (2) *limiting numbers,* those which have no number immediately preceding them, from which they are formed by the addition of unity; for example ω, $\omega^2+\omega$, $\omega^\omega+\omega^2$ are limiting numbers.

Any particular number of the second class can be denoted by a finite number of symbols, but there is no upper limit to the number of symbols required to denote such numbers.

Cantor has further postulated the existence of a number Ω which comes after all the numbers of the second class, and is the first number of a new set which is called the third class. The number Ω cannot be obtained as the number which succeeds a simple sequence, by means of the second principle of generation; for every number which can be so obtained is itself a number of the second class. This number Ω can be obtained only by means of a third principle of generation, which postulates the existence of a new number coming after all the numbers of the complex formed by the application of the first and second principles of generation. The validity of the postulation of the existence of the number Ω, and of the higher numbers of the third class will be discussed in Chapter III.

62. A fundamental property of the numbers of the second class may be expressed as follows:—

Let $P_1, P_2, P_3, \dots P_n, \dots P_\omega, P_{\omega+1}, \dots$ be an infinite set of points such that either (1) there is a last point P_β, where β is some number of the second class, or (2) there is no last point, but every index occurs which is less than some limiting number γ of the second class, whereas the index γ itself does not occur; the set of points is then enumerable.

The sets

$$\begin{array}{llll} P_1, & P_2, & \dots P_n, \dots \\ P_\omega, & P_{\omega+1}, & \dots P_{\omega+n}, \dots \\ P_{\omega.2}, & P_{\omega.2+1}, & \dots P_{\omega.2+n}, \dots \\ P_{\omega.3}, & P_{\omega.3+1}, & \dots \\ \dots\dots & & \\ P_{\omega.r}, & P_{\omega.r+1}, & \dots \\ \dots\dots & & \end{array}$$

where every index less than ω^2 occurs, form an enumerable aggregate of enumerable sets of points; and this has been shewn in § 54, to be itself an enumerable set. Now consider the sets

$$\begin{array}{lllll} P_1, P_2, P_3, & \dots P_\omega, & P_{\omega+1}, \dots P_{\omega.2}, \dots P_{\omega.3}, \dots \\ P_{\omega^2}, & P_{\omega^2+1}, & \dots P_{\omega^2+\omega}, & P_{\omega^2+\omega+1}, \dots \\ P_{\omega^2.2}, & P_{\omega^2.2+1}, & \dots P_{\omega^2.2+\omega}, \dots \\ P_{\omega^2.3}, & P_{\omega^2.3+1}, & \dots P_{\omega^2.3+\omega}, \dots \\ \dots\dots \\ \dots\dots \end{array}$$

such that in the first set there is every index less than ω^2, in the second, every index less than $\omega^2.2$, and so on. Each of these sets is enumerable, and there is an enumerable set of sets; hence the whole set, which contains every index less than ω^3, is enumerable. In this manner it can be shewn that, if every index less than ω^n occurs, the set is enumerable. If the theorem holds for sets which contain every index less than $\beta_1, \beta_2, \beta_3, \dots$, then it holds for a set which contains every index less than β, the limiting number of the sequence $\beta_1, \beta_2, \beta_3, \dots$. For the points with indices less than β_1, with indices $\geqq \beta_1$ and $< \beta_2$, with indices $\geqq \beta_2$ and $< \beta_3$, &c. form an enumerable sequence of enumerable sets; therefore by the theorem of § 54, the whole set with indices $< \beta$ is enumerable. Since the theorem holds for $\beta_1 = \omega$, $\beta_2 = \omega^2$, $\beta_3 = \omega^3, \dots$ it holds for $\beta = \omega^\omega$. By continual application of this method, since any number can be reached by means of the two principles of generation, and since every number is either a limiting number, or is obtained from one by adding a finite number, we see that the general theorem holds.

It will now be shewn, conversely, that *if a set of points $P_1, P_2, \dots P_n, \dots P_\omega, \dots P_\beta, \dots$ is enumerable, there must be some definite number γ of the first or of the second class, such that γ does not occur among the indices of the points, and such that every number less than γ does so occur.*

In case γ is a limiting number, there is no last point of the set; but if γ is not a limiting number, there is a last point, viz. the one of which the index is the number immediately preceding γ.

To prove the theorem, we observe that, since the given set of points is enumerable, it may be placed in correspondence with a set of points Q_1, $Q_2, \dots Q_n, \dots$ in which all the indices are numbers of the first class. Let us suppose, that if possible, no number γ exists, and let P_{α_1} be the point of $\{P\}$ which corresponds to the point Q_1 of $\{Q\}$. Let Q_{p_1} be the point of $\{Q\}$ of smallest index, such that the corresponding point of $\{P\}$ has an index which is $> \alpha_1$; denote this index by α_2. Then let Q_{p_2} be that point of $\{Q\}$, of smallest index, such that the corresponding point of $\{P\}$ has an index $> \alpha_2$; denote this index by α_3. Proceeding in this manner, we have a set of points $Q_1, Q_{p_1}, Q_{p_2}, \dots Q_{p_n}, \dots$ corresponding in order to a set of points $P_{\alpha_1}, P_{\alpha_2}, P_{\alpha_3}, \dots P_{\alpha_n}, \dots$ where $\alpha_1 < \alpha_2 < \alpha_3 \dots < \alpha_n < \dots$. There exists a number α of the second class, which is the limit of the sequence $\alpha_1, \alpha_2, \dots \alpha_n, \dots$; and by hypothesis there exists a point P_α, which has α for index. Now the set $\{Q\}$ can contain no point which corresponds to P_α, because each point Q_n corresponds to a point of $\{P\}$ with an index less than α, and thus there is a contradiction in the hypothesis that α occurs amongst the indices of the points of $\{P\}$. Hence there exist numbers of the second class which do not occur as indices in the set $\{P\}$, and these numbers form a set which is a part of the aggregate of numbers of the second class. In this set there must be a lowest number γ, and this number γ is the first which does not occur amongst the indices of the set $\{P\}$. That every part of the aggregate of numbers of the first and second classes, has a lowest number, will be shewn in Chapter III., to be a consequence of the structure of the ordered aggregate.

EXAMPLES.

1. On a straight line AB, let us denote by $P_1, P_2, P_3, \dots$, those points at which the ratio AB/PB has the values 1, 2, 3, ... The point P_1 coincides with A, and the point B can only be represented by P_ω. Now take any one of the segments P_rP_{r+1}; this may for convenience be represented on an enlarged scale. Denote by $Q_{r1}, Q_{r2}, Q_{r3}, \dots$, the points on P_rP_{r+1}, at which P_rP_{r+1}/QP_{r+1} takes the values 1, 2, 3, ...; thus P_{r+1} can only be represented by $Q_{r\omega}$. Supposing this to have been done with every segment P_rP_{r+1} of AB, let us imagine all the points Q to be marked on AB, and to be numbered from left to right.

In P_1P_2, we shall have $1, 2, 3, \dots \omega$,

in P_2P_3 there will be $\omega+1, \omega+2, \dots \omega . 2$,

and in P_3P_4 $\omega . 2+1, \omega . 2+2, \dots \omega . 3$;

the point B can be represented only by ω^2. If now we proceed to take each segment $Q_{rs}Q_{r,s+1}$, and to divide it in a similar manner, at points R for which $Q_{rs}Q_{r,s+1}/RQ_{r,s+1}$

has the values 1, 2, 3, ..., and then imagine all the points R obtained in every such segment $Q_{rs}Q_{r,s+1}$ to be marked on AB, and numbered as before, from left to right, it will be seen that all the numbers $\omega^2 p+\omega q+r$ will be required, and that the point B can be represented by ω^3. The points P_1, P_2, ... P_ω will have for their ordinal numbers 1, ω^2, $\omega^2 . 2$, $\omega^2 . 3, \ldots \omega^3$; the point Q_{rs} will be numbered $\omega^2 . r+\omega . s$; the finite numbers are all used up in the first sub-segment of AB. By proceeding to further subdivision, we may exhibit on AB, the ordinal numbers $\omega^n p_n+\omega^{n-1}p_{n-1}+\ldots+p_0$, and the point B will then be represented by ω^{n+1}.

2. The properties of the integral numbers in relation to their prime factors may be employed to rearrange the series 1, 2, 3, ... n, ..., so that the numbers may be made to correspond with a series of ordinal numbers of the first and second classes.

First take the primes 1, 2, 3, 5, 7, 11, ... ; these correspond with the numbers of the first class 1, 2, 3, ... n, ... Then take the squares of the primes, omitting unity; we thus have $2^2, 3^2, 5^2, 7^2, 11^2, \ldots$, corresponding to $\omega, \omega+1, \omega+2, \ldots \omega+n, \ldots$.

We then take the cubes of the primes,

$$2^3, 3^3, 5^3, 7^3, 11^3, \ldots, \text{ corresponding to } \omega . 2, \omega . 2+1, \ldots \omega . 2+n, \ldots,$$

and in general, 2^{r+1}, 3^{r+1}, 5^{r+1}, ..., corresponding to $\omega . r$, $\omega . r+1, \ldots \omega . r+n, \ldots$. We may then take the numbers ab which consist of the product of two prime factors; these arranged in ascending order correspond to ω^2, $\omega^2+1, \ldots \omega^2+n, \ldots$. Next take the numbers a^2b^2, which consist of the squares of the last set; these correspond to $\omega^2+\omega$, $\omega^2+\omega+1, \ldots$. We then take the successive sets of numbers of the forms a^3b^3, $a^4b^4, \ldots$; we thus obtain the numbers which may be taken to correspond with

$$\omega^2+\omega . 2,\ \omega^2+\omega . 2+1, \ldots \omega^2+\omega . 3, \ldots \omega^2+\omega . p+q, \ldots,$$

all of which are less than $\omega^2 . 2$. The sets of numbers of the forms

$$a^2b,\ (a^2b)^2, \ldots (a^2b)^n, \ldots a^3b,\ (a^3b)^2, \ldots (a^3b)^n, \ldots a^4b,\ (a^4b)^2, \ldots,$$

may then be taken. Afterwards, we may proceed with the numbers which contain three different prime factors, and so on. It is clear that this mode of rearranging the integral numbers in their natural order, so that they correspond in the new order with ordinal numbers of the first and second classes, admits of great variety. In every case, there will be some lowest number of the second class, which is not employed in the correspondence established.

THE TRANSFINITE DERIVATIVES OF A SET OF POINTS.

63. If G denotes a set of points in the interval (a, b), it has been shewn that the derivatives $G^{(1)}, G^{(2)}, \ldots G^{(n)}$ are all closed sets, and that all the points of any one of these sets, after the first, are contained in the preceding set. If G is of the second species, then $G^{(n)}$ exists for all values of n; and in this case the set $D(G^{(1)}, G^{(2)}, \ldots G^{(n)} \ldots)$, which contains points belonging to every $G^{(n)}$, is denoted by $G^{(\omega)}$, where ω is the first transfinite number. It will be shewn that $G^{(\omega)}$ contains one point at least, and is a closed set. It is defined to be the derivative* of G of order ω.

If p_1 is a point of $G^{(1)}$, p_2 a point of $G^{(2)}, \ldots p_n$ a point of $G^{(n)}$, &c., the points $p_1, p_2, \ldots p_n, \ldots$ form a set $\{p_n\}$ which has at least one limiting point p.

* See Cantor, *Math. Annalen*, vol. XVII.

This point p belongs to $G^{(n)}$ whatever value n has, because all except a finite number of the set $\{p_n\}$ are points of $G^{(n)}$; and therefore p is a point of G_ω. Let $q_1, q_2, \ldots q_n, \ldots$ be a sequence of points of G_ω, in case G_ω contains more than a finite number of points; and suppose this sequence to have the limiting point q. Then since all the points $q_1, q_2, \ldots q_n, \ldots$ are points of the closed set $G^{(n)}$, the limiting point q is a point of $G^{(n)}$; and this holds for every value of n, hence q is a point of $G^{(\omega)}$, and therefore $G^{(\omega)}$ is a closed set.

We can proceed to form the derivatives of $G^{(\omega)}$ in a similar manner to that in which the derivatives $G^{(1)}, G^{(2)}, \ldots$ of G were formed. These successive derivatives are denoted by $G^{(\omega+1)}, G^{(\omega+2)}, \ldots G^{(\omega+n)}, \ldots$ and are regarded as the derivatives of G of the transfinite orders $\omega+1, \omega+2, \ldots \omega+n, \ldots$. They have the same properties as the derivatives of finite order, viz. that all the points of each are points of $G^{(1)}$, and that all the points of any one of them are points of the preceding ones.

It may happen that one of the derivatives $G^{(\omega+n)}$ contains no points; then the process of forming derivatives has come to an end, the last one being $G^{(\omega+n-1)}$. If this is not the case, a repetition of the above reasoning shews that the set $D(G^{(\omega+1)}, G^{(\omega+2)}, \ldots G^{(\omega+n)}, \ldots)$ contains at least one point, and is a closed set; this set is denoted by $G^{(\omega.2)}$, and is defined to be the derivative of G of order $\omega.2$. In the same manner we can proceed to form further derivatives, whose orders are numbers of the second class.

In general, if $\alpha_1, \alpha_2, \alpha_3, \ldots \alpha_n, \ldots$ denote a sequence of numbers of the second class whose limiting number is β, the same reasoning as before shews that, if all the derivatives $G^{(\alpha_1)}, G^{(\alpha_2)}, \ldots G^{(\alpha_n)}, \ldots$ exist, then the set $D(G^{(\alpha_1)}, G^{(\alpha_2)}, \ldots G^{(\alpha_n)}, \ldots)$ contains at least one point, and is a closed set. This is denoted by $G^{(\beta)}$, and is defined to be the derivative of G of order β.

If we form the successive derivatives of the set G, whose orders are the numbers of the first and second classes, it may happen that there is a first number γ, of the first or second class, for which $G^{(\gamma)} \equiv 0$; but this number γ cannot be a limiting number of the second class.

It may, however, happen that no number γ, of the first or second class, exists for which $G^{(\gamma)} \equiv 0$, so that derivatives of G exist of orders corresponding to all the numbers of the first and second classes. It will be shewn in § 73, that if $G^{(\gamma)}$ does not vanish, for some number γ, of the first or of the second class, then there necessarily exists a number β, of the first or second class, such that $G^{(\beta)} = G^{(\beta+1)} = G^{(\beta+2)} \ldots$. This set $G^{(\beta)}$ is a perfect set, and it is frequently denoted by $G^{(\Omega)}$, where Ω is the first transfinite number of the third class. The notation $G^{(\Omega)}$ may however be employed, independently of the acceptance of the theory of numbers of the third class.

Conversely, if $G^{(\Omega)}$ does not exist, $G^{(\gamma)}$ must first vanish for some number γ of the first or second class, which number cannot be a limiting number.

If $G_1, G_2, G_3, \ldots G_n, \ldots$ be any endless sequence of sets of points, such that each set G_n is contained in the preceding one G_{n-1}, then the set $D(G_1, G_2, \ldots G_n, \ldots)$, if it exists, consists of those points each of which belongs to G_n for every value of n, and this set may be denoted by G_ω. Commencing with G_ω, a new sequence of sets $G_\omega, G_{\omega+1}, G_{\omega+2}, \ldots G_{\omega+n}, \ldots$ may be considered, each one being a part of the preceding one; the set of points each of which belongs to all the sets of this sequence is $D(G_\omega, G_{\omega+1}, \ldots G_{\omega+n}, \ldots)$, and may, when it exists, be denoted by $G_{\omega.2}$. In this manner further sets may be formed, requiring as indices, higher numbers of the second class. An example illustrating the fact that G_ω does not necessarily exist is

$$G_n = \left(\frac{1}{n}, \frac{1}{n+1}, \frac{1}{n+2}, \ldots\right).$$

The case of the transfinite derivatives of a given set G, considered above, is a special case of a sequence of sets each one of which contains the next one.

EXAMPLES.

1*. Let G denote the enumerable set of points, each one of which is given by

$$\frac{1}{2^n} + \frac{1}{2^{n+m_1}} + \frac{1}{2^{n+m_1+m_2}} + \ldots + \frac{1}{2^{n+m_1+m_2+\ldots+m_n}},$$

when n has all positive integral values, excluding zero, and $m_1, m_2, \ldots m_n$ have all positive integral values including zero, independently of one another.

It is easily seen that in $G^{(n)}$, the points $\frac{1}{2^n}, \frac{1}{2^{n+1}}, \ldots$ all occur, and hence that $G^{(\omega)}$ exists, and consists of the single point zero.

2*. Let G denote the enumerable set of points, each one of which is given by

$$\frac{1}{2^{m_1}} + \frac{1}{2^{m_1+m_2}} + \ldots + \frac{1}{2^{m_1+m_2+\ldots+m_n}} + \frac{1}{2^{m_1+m_2+\ldots+m_n+p}} + \frac{1}{2^{m_1+m_2+\ldots+m_n+p+q_1}}$$
$$+ \frac{1}{2^{m_1+m_2+\ldots+m_n+p+q_1+q_2}} + \ldots + \frac{1}{2^{m_1+m_2+\ldots+m_n+p+q_1+q_2+\ldots+q_p}},$$

where $m_1, m_2, \ldots m_n, p, q_1, q_2, \ldots q_p$ have all positive integral values, including zero. In this case $G^{(\omega+n)}$ consists of the single point zero.

3*. Let G denote the enumerable set of points, each one of which is given by

$$\frac{1}{2^n} + \frac{1}{2^{n+m_1}} + \frac{1}{2^{n+m_1+m_2}} + \ldots + \frac{1}{2^{n+m_1+\ldots+m_n}} + \frac{1}{2^{n+m_1+\ldots+m_n+p}}$$
$$+ \frac{1}{2^{n+m_1+\ldots+m_n+p+q_1}} + \ldots + \frac{1}{2^{n+m_1+m_2+\ldots+m_n+p+q_1+q_2+\ldots+q_p}},$$

where $n, m_1, m_2, \ldots m_n, p, q_1, \ldots q_p$, have all positive integral values. In this case $G^{(\omega\cdot 2)}$ exists, and consists of the single point zero.

SETS OF INTERVALS.

64. The properties of a set of intervals, which intervals are assigned in any manner, are closely connected with the properties of sets of points, and will therefore be considered here in some detail.

* These examples were given by Mittag-Leffler, *Acta Math.* vol. IV, p. 58.

If two intervals have only an end-point of each in common, they are said to *abut on one another*; and if the two intervals have more than one point in common, they are said to *overlap one another*.

*Every set of intervals, which is such that no two of the intervals overlap, is an enumerable aggregate**.

First, suppose the set of non-overlapping intervals to lie in the finite segment (a, b); and choose a sequence $\epsilon_1, \epsilon_2, \ldots \epsilon_n, \ldots$ of positive numbers converging to the limit zero. The number of intervals of the given set which are of length greater than, or equal to ϵ_n, is finite, since it cannot exceed $\frac{1}{\epsilon_n}(b-a)$. We can now arrange the intervals in order of magnitude, taking first those which are $\geqq \epsilon_1$, then those which are $< \epsilon_1$ and $\geqq \epsilon_2$, and so on, there being only a finite number in each set. Therefore, since the set of intervals can be arranged as a simply infinite aggregate, it is an enumerable set.

Next, suppose that the intervals are on an unlimited straight line in which the position of any point is denoted by x. If we consider the correspondence given by $x' = \frac{x}{\sqrt{x^2+1}}$, where the radical has always the positive sign, the unlimited straight line corresponds to the segment $(-1, 1)$, in which the point x' lies. The intervals of the given set correspond uniquely to intervals of a non-overlapping set in the segment $(-1, +1)$, and this latter set is enumerable; hence the given set is so also.

The theorem can be generalized so as to apply to the case of detached portions of space of two, three, or any number of dimensions. If within a finite portion of such space, there be a set of portions no two of which overlap one another, though they may have portions of their boundaries in common, the set of such portions is enumerably infinite if it be not finite.

The theorem is proved, as in the case of intervals in a one-dimensional space, from the consideration that there can only be a finite number of the portions of volume† greater than, or equal to ϵ_n.

Since the points of unbounded space, say of three dimensions, can be made to correspond with the points of a finite portion of space, by means of the transformation

$$x' = \frac{x}{\sqrt{x^2+1}}, \quad y' = \frac{y}{\sqrt{y^2+1}}, \quad z' = \frac{z}{\sqrt{z^2+1}},$$

the restriction that all the portions must be contained in a finite domain can be removed.

* Cantor, *Math. Annalen*, vol. xx.

† More generally "measure," see § 81, below.

65. The theorem which is given in § 64, can now be applied to prove that *every isolated set of points is enumerable**.

Let P be a point of such a set. Since in a sufficiently small neighbourhood of P, no other points of the set occur, take such a neighbourhood of length ρ, and conceive such neighbourhoods to be chosen for every point of the set; we now have a set of non-overlapping intervals which is enumerable, and therefore the isolated set of points is also enumerable.

It has been shewn that any set of points G is made up of an isolated aggregate, and of one which is a divisor of G'. It follows that, if the derivative G' is enumerable, so also is G; but the converse does not hold.

Every set of points which is of the first species is enumerable. For, if s be its order, $G^{(s)}$ contains only a finite number of points; hence $G^{(s-1)}$ is enumerable; and therefore also $G^{(s-2)}$, $G^{(s-3)}$, ... G are all enumerable sets.

A set of points of the second species is enumerable if one of its derivatives be so. If any set G is not enumerable none of its derivatives is so.

66. Let us consider a given set of overlapping intervals contained in the finite segment (a, b); it will be shewn that the given set can be replaced by a set of non-overlapping intervals which is such, that every point which is interior to any interval of either set is interior also to some interval of the other set.

Taking any point $P(x)$ which is an interior point of one or more intervals of the given set, the points x' of the segment (x, b) can be divided into two classes, those for which every point interior to the segment (x, x') is an interior point of some interval of the given set, and those for which this is not the case. This section of the numbers in the segment (x, b) defines a single point x', such that (x, x') is the greatest segment on the right of x which has every interior point of it also an interior point of the given set of intervals. Similarly a definite segment (x'', x) on the left of x, can be found, which has the corresponding property. Therefore the interval (x'', x') is one of the required intervals. If now we take any point in either of the parts of (a, b) complementary to (x'', x'), which is an interior point of an interval of the given set, we may proceed as before to construct an interval of the required set which contains that point; and so on, until we have a set of non-overlapping intervals which contain, as interior points, every point that is interior to any interval of the given set. It has thus been proved that:—

Every set of intervals contained in a finite segment can be replaced by a set of non-overlapping intervals of which the interior points are the same as those of the given set.

* Cantor, *Math. Annalen*, vol. XXI.

The new set may be spoken of as the set of non-overlapping open intervals equivalent to the given set of open intervals.

An *open interval* PQ is defined as in § 40, to consist of the aggregate of points interior to PQ, excluding the end-points P, Q.

The properties of any set of open intervals in a finite segment thus depend upon those of a non-overlapping set of such intervals, and we proceed to the consideration of the latter.

Every point of (a, b) which is not interior to an interval of the non-overlapping set is either

(1) a common end-point of two intervals of the given set; or

(2) a point interior to, or at an end of, an interval not belonging to the given set, this interval containing no point which is interior to any interval of the set; or

(3) a limiting point, on both sides, of end-points of intervals of the set; or

(4) an end-point of an interval of the given set, and also a limiting point, on one side, of end-points of intervals of the given set.

If either a or b is an end-point of an interval, we reckon that point as belonging to the points (1).

The points described in (2) or (3) may be described as *external points* of the given set; and if a or b is a limiting point of end-points, it will be reckoned as an external point.

The points described in (4) may be spoken of as *semi-external** points.

The following theorem will now be established:—

Those points of the segment (a, b), *which are not points of a given set of non-overlapping open intervals, form a closed set of points.*

The closed set includes all the end-points of the given set of intervals, and all the external points.

To prove this theorem, we observe that no limiting point of the set of points, complementary to the set of open intervals, can be interior to an interval of the given set. For if P be such an interior point, a neighbourhood of P exists, viz. the interval in which it is contained, within which there are no points of the set; and thus P cannot be a limiting point of the set. All the limiting points of the set must therefore belong to the set itself, which is consequently closed.

The closed set which is complementary to a set of non-overlapping *open* intervals contains all the end-points of the intervals, all those points which, not being end-points, are limiting points on both sides of end-points of

* This term is due to W. H. Young, *Proc. Lond. Math. Soc.* vol. xxxv, p. 250.

intervals, and also the points interior to the complementary intervals, in case such complementary intervals exist.

In case there are no complementary intervals, then the closed set of points defined as the set complementary to a given set of open intervals, is a non-dense closed set.

67. It will now be shewn that*, *unless a given set of non-overlapping intervals is a finite set, there must be at least one external or semi-external point*; in other words the whole interval (a, b) cannot be filled up by an indefinitely great number of non-overlapping intervals, each one of which abuts on the next, without leaving at least one point over, which is neither interior to an interval nor is an end-point of two intervals, the points a, b being regarded as end-points of two intervals if they are end-points of one interval of the given set.

If there be any complementary intervals, then the points of these intervals are all external points, and we therefore need only consider the case in which no such complementary intervals exist. We observe that, when the number of intervals is not finite, their end-points must have at least one limiting point P. Now this point P cannot be interior to one of the given intervals; for, if it were so, it would have a neighbourhood, viz. the interval to which it is interior, within which are no end-points. Neither can P be a common end-point of two intervals; for it would then have a neighbourhood on the right, and also one on the left, within which there is no end-point except P itself. The point P must consequently either be an external point, *i.e.* one which is not an end-point but is a limiting point, on both sides, of end-points; or else it must be an end-point of one interval, and a limiting point, on one side, of end-points. If a, or b is not an end-point, it is regarded as an external point. It will subsequently be shewn that the external and semi-external points form a set which may be either finite, or of cardinal number a, or of cardinal number c.

EXAMPLES.

1†. In the interval (0, 1) take the intervals $(0, \frac{1}{4})$, $(\frac{1}{4}, \frac{3}{8}) \ldots \left(\frac{2^{n-1}-1}{2^n}, \frac{2^n-1}{2^{n+1}}\right) \ldots$, and also the intervals obtained by reflecting these intervals in the point $\frac{1}{2}$. The point $\frac{1}{2}$ is external to all the intervals, and yet the limiting sum of the intervals is equal to 1, the length of the whole interval (0, 1) in which the enumerable set of intervals is contained.

If instead of reflecting the intervals in the point $\frac{1}{2}$, we take the interval $(\frac{1}{2}, 1)$, the point $\frac{1}{2}$ is now a semi-external point, and the limiting sum of the intervals is the same as before.

2†. Take the set $(\frac{1}{2}, 1)$, $(0, \frac{1}{4}) \ldots \left(\frac{2^{n-1}-1}{2^n}, \frac{2^n-1}{2^{n+1}}\right) \ldots$ of intervals, and divide each

* This theorem was given by W. H. Young, *Proc. Lond. Math. Soc.* vol. xxxv, p. 251.

† See W. H. Young, *Proc. Lond. Math. Soc.* vol. xxxv, pp. 249—251.

interval into a set of sub-intervals similar to the whole. We now have a new enumerable set of intervals which has no external points, but of which the semi-external points form an enumerable set $\frac{3}{4}, \frac{1}{2}, \frac{1}{8}, \frac{5}{16}, \ldots$

68. *If a set of intervals in (a, b) is such that every point of (a, b) is an interior point of at least one interval (the end-points a, b being each an end-point of at least one interval), then a finite number of the intervals can be selected which has the same property as the whole set.*

This theorem, which is known as the Heine-Borel theorem*, is of considerable importance in the theory of functions, and may be proved as follows:—

Denoting the points a, b, by A, B, we may select an interval Aq_1 which has A as end-point; then select an interval p_2q_2, of which q_1 is an interior point; then p_3q_3, of which q_2 is an interior point, and so on; and consider

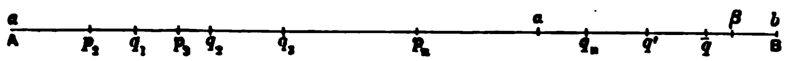


FIG. 2.

the points $q_1, q_2, q_3, \ldots$ thus constructed. If one of these points q_n coincides with B, the finite set of intervals $Aq_1, p_2q_2, \ldots p_nq_n$ required has been found. If q_n does not coincide with B for any value of n, then the infinite set of points $q_1, q_2, \ldots q_n, \ldots$ has a limiting point q' which is on the right of all of them. Let us now suppose that it is impossible to select a finite number of the intervals in the manner described, so that the end-point of the last is at B. Then whatever particular selection of intervals we make as above, we obtain a point q' on the right of the intervals, the position of q' depending on the selection made. The set of points $\{q'\}$ which has thus been obtained, has either (1) a limiting point $\bar{q}$ on the right of all the points of $\{q'\}$, or (2) an extreme point $\bar{q}$, belonging to the set, on the right of all the other points of the set: and in either case $\bar{q}$ may or may not coincide with B. In case (1), the point $\bar{q}$ is interior to an interval $\alpha\beta$ of the given set, or else is an end-point of such, β, $\bar{q}$ then coinciding with B. We can now choose a set of intervals $Aq_1, p_2q_2, \ldots p_nq_n, \ldots$ for which the limiting point q' lies within $\alpha\bar{q}$, since $\bar{q}$ is the limiting point of $\{q'\}$; and then only a finite number of the points

$$q_1, q_2, \ldots q_n, \ldots$$

lie outside $\alpha q'$. Let then q_n be the first of them which lies inside $\alpha q'$, and consider the set of intervals $Aq_1, p_2q_2, \ldots p_nq_n, \alpha\beta$. If $\bar{q}$ and β coincide, and are therefore both coincident with B, we have here a finite set of intervals such as the theorem requires, and this is contrary to the hypothesis made that no such finite set exists. If, on the other hand, $\bar{q}$ and β do not coincide,

* Borel, *Ann. de l'Éc. norm.* (3) XII. p. 51. See also Borel's *Leçons sur la théorie des fonctions*, p. 42.

it is impossible that $\bar{q}$ should be the limit on the right of the limiting points of all the possible sets $q_1, q_2, \dots q_n, q_{n+1}, \dots$; for we may take q_{n+1} to coincide with β, which is itself on the right of $\bar{q}$. We have therefore again a contradiction. In case (2), there is one set of intervals

$$Aq_1,\ p_2q_2,\ \dots\ p_nq_n,\ \dots$$

such that $\bar{q}$ is the limiting point of $q_1, q_2, \dots q_n, \dots$; and the position of q' for any other set is on the left of $\bar{q}$. Now $\bar{q}$ is interior to an interval $\alpha\beta$ of the given set, and only a finite number of the points $q_1, q_2, \dots q_n, \dots$ is on the left of α. Let q_n be the first which is inside $\alpha\beta$; then the set

$$Aq_1,\ p_2q_2,\ \dots\ p_nq_n,\ \alpha\beta$$

is a finite set such as the theorem requires, in case β coincides with B. But if β does not coincide with B, it is part of an infinite set for which the limit of $q_1, q_2, \dots q_n, \beta, \dots$ is on the right of $\bar{q}$, which is contrary to the supposition that $\bar{q}$ has the extreme position on the right for all points of the set $\{q'\}$. There is therefore, as in the other case, a contradiction in supposing that B cannot be reached after taking a finite number of intervals. It will be observed, that the set of intervals contemplated in the theorem is not necessarily enumerable.

The theorem may be stated in a somewhat different form, in which it is capable of being proved in a simple manner.

Let us suppose that with each point of (a, b) is associated an interval of which the point is an interior point, the intervals associated with a, or b, extending beyond (a, b). Let the associated interval be called the *proper interval* of the point. Further, let any interval be provisionally called a *suitable interval*, when it is included in the proper interval of some point within or upon the boundary of itself.

The theorem may then be stated*, that *the interval (a, b) can be divided into a finite number of suitable intervals.*

For, let the interval (a, b) be halved; if one of the halves is not suitable, let it be halved; and so on. This halving process, which is to be applied to every interval not already suitable, indefinitely, will terminate after a finite number of steps. For otherwise, let us consider an indefinitely continued sequence of intervals, each half of its predecessor, and no one of them suitable. These intervals determine a single point within or upon the boundary of every one of them. Let us consider the proper interval of this point; the sequence of intervals of which the point is the limiting point, will, from and after some fixed member of the sequence, all lie within this proper interval of the point. This is contrary to the hypothesis that the sequence of unsuitable intervals is indefinitely continued.

* The theorem stated in this form is really contained in Goursat's proof of Cauchy's theorem; see *Trans. Amer. Math. Soc.*, vol. I, p. 15.

69. The Heine-Borel theorem can be extended to the case of sets in two, three, or any number of dimensions. In the case of a set in two dimensions, we may suppose for simplicity that the set of areas is contained in a rectangular area $ABDC$. We suppose that there exists a set of closed areas, which may be, for example, all circles, or all rectangles, such that each point

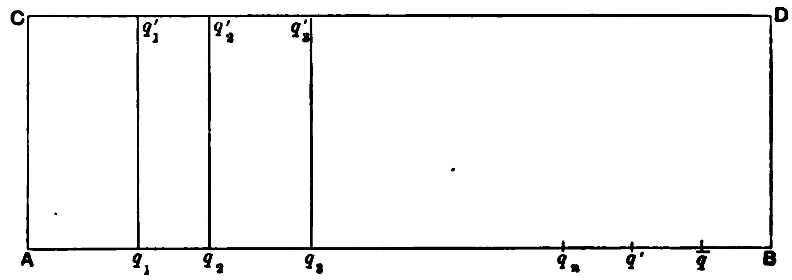


FIG. 3.

inside $ABDC$ is interior to one at least of the areas, and that each point on the boundary of $ABDC$ is interior to the straight boundary of at least one such area. Since all the points of AC are interior, in the sense explained, to areas of the given set, and since these areas are bounded by intervals on AC, the Heine-Borel theorem proved above, shews that a finite number of areas can be selected such that all the points on AC are interior points of them. A straight line q_1q_1' can then be found such that all the points interior to the area $Aq_1q_1'C$ are interior points of the areas which have been already selected. Next we see in a similar manner that all the points of q_1q_1' may be enclosed as interior points of a properly selected finite set of the given areas; we then see that a point q_2 exists to the right of q_1, such that the areas already determined enclose all the interior points of $q_1q_1'q_2'q_2$ as internal points; and proceeding in this manner, we obtain a set of intervals Aq_1, q_1q_2, q_2q_3, ... on AB. Now the point B must be reached at the end of a finite number of stages of this process; for we may shew by precisely the same reasoning as before, that it is impossible but that the point B be reached at a finite stage of the process of taking in new finite sets selected from the given set of areas.

Assuming the truth of the theorem for sets of areas in two dimensions, it may be extended, in an analogous manner, to a three-dimensional space, and so on to spaces of any number of dimensions.

It is clear that Goursat's form of the Heine-Borel theorem may be proved in the case of sets of any number of dimensions, exactly as in the proof given in § 68, for the case of linear sets. The division of a rectangular cell of n dimensions into 2^n equal rectangular cells, will replace the process of halving applicable to linear intervals.

70. The following theorem will now be proved:—

If any unenumerable set of overlapping intervals in (a, b) be given, then an enumerable set can be selected out of the intervals of the given set, of which the interior points are the same as those of the given set.

It has been shewn in § 66, that the given set can be replaced by a non-overlapping set of intervals with the same interior points. An interval of this second set is however not in general an interval of the given set.

Let PQ be an interval of the equivalent non-overlapping set; then every internal point of PQ is an internal point of one interval at least of the given set. The point P is either an end-point of some interval Pp of the given set, or else it is a limiting point of end-points of an infinite number of intervals of the given set. In the latter case we can choose an enumerable sequence $P_1p_1, P_2p_2, P_3p_3, \ldots$ of intervals of the given set such that P is the limiting point of the sequence of points $P_1, P_2, \ldots P_n, \ldots$. Similarly, unless Q is an end-point of an interval qQ of the given set, it is the limiting point of a sequence $Q_1, Q_2, \ldots Q_n, \ldots$ of end-points of intervals $q_1Q_1, q_2Q_2, \ldots q_nQ_n, \ldots$ of the given set. Consider the intervals $P_1Q_1, P_2Q_2, \ldots P_nQ_n, \ldots$, where $P_1, P_2, \ldots$ may be taken all to coincide with P in case the interval Pp exists, a similar convention being made as regards Q. Since every point of P_1Q_1 is interior to some interval of the given set, therefore in accordance with the Heine-Borel theorem, a finite number of intervals of the given set can be selected so that every point of P_1Q_1 is interior to one at least of them. Let a similar selection of a finite set of intervals be made for each of the intervals $P_2Q_2, P_3Q_3, \ldots P_nQ_n, \ldots$; we have then altogether an enumerable set of finite sets of intervals. The totality of these intervals forms a finite, or an enumerable, set of intervals selected from the given set, which contains every point in the interior of PQ as an interior point. Applying the same process to each interval PQ of the equivalent non-overlapping set, and remembering both, that the intervals PQ form a finite or enumerable set, and that an enumerable set of finite or enumerable sets is itself enumerable, we derive the conclusion that an enumerable set of intervals can be selected from the given set such that the internal points are identical with those of the given set.

71. The Heine-Borel theorem can be extended to the case where the points, which are to be internal to a finite number of intervals selected from a given set, are a given closed set of points, instead of the whole set of points of the segment in which the intervals lie.

Let a given set of intervals in (a, b) be such that every point of a given closed set of points is interior to one at least of the given intervals. Consider the set of non-overlapping intervals equivalent to the given set; this set must be finite; for, if not, it has at least one external or semi-external point P which is a limiting point of the end-points of the intervals. Then any arbitrarily small neighbourhood of P contains an indefinitely great number of end-points of intervals, and therefore also of points of the given closed set; and P would therefore be a limiting point of the closed set, but this is impossible, as P does not belong to that set. Let pq be one of this finite number of intervals of the equivalent non-overlapping set; then the part of the given closed set of points which is in pq is itself closed. In pq take an interval $p'q'$ which contains this closed part of the given set of points in its interior: then by the Heine-Borel theorem a finite number of intervals can be selected from the given set of intervals which contains every point of $p'q'$ as an internal point; and therefore contains the part of the closed set of points which is interior to $p'q'$. Applying this process to each of the finite number of intervals pq we have the following theorem*:—

Having given a closed set of points in (a, b), *and a set of intervals such that each point of the closed set is interior to one interval at least of the set, a finite number of intervals can be selected from the given set which is also such that every point of the closed set of points is interior to one at least of these intervals.*

The proof of Goursat's form of the Heine-Borel theorem can be modified so as to apply to this case. We have only to neglect, in the proof of § 68, those sub-intervals which do not contain any of the points of the given closed set.

NON-DENSE CLOSED AND PERFECT SETS.

72. It has been shewn in § 66, that if an infinite number of non-overlapping intervals be contained in (a, b), the set of points which is complementary to the internal points of the intervals forms a closed set. If no interval whatever can be found in (a, b) every point of which belongs to the closed set, the given set of intervals is everywhere-dense, and the closed set is in no interval everywhere-dense, and is therefore said to be non-dense in (a, b). We shall now prove the converse theorem† that:—

Every non-dense closed set of points consists of the end-points of a set of non-overlapping intervals which is everywhere-dense in the domain, and of the limiting points of such end-points.

* See W. H. Young, *Proc. Lond. Math. Soc.* vol. XXXV, p. 387; also Borel, *Comptes Rendus*, January, 1905. For a further extension of the theorem, see W. H. Young, *Messenger of Math.* vol. XXXIII, p. 129, and also *Proc. Lond. Math. Soc.* Ser. 2, vol. II, p. 67.

† This relation between everywhere-dense sets of intervals and closed sets was discovered by Du Bois Reymond and by Harnack. See Du Bois Reymond's *Allgemeine Functionentheorie* (1882), p. 188; also *Math. Annalen*, vol. XVI, p. 128, where everywhere-dense sets of intervals are introduced. See also Harnack, *Math. Annalen*, vol. XIX, p. 239, and Bendixson, *Acta Math.* vol. II, p. 416, and *Öfv. af. Svensk. Vet. Forh.* vol. XXXIX, 2, p. 31. Proofs of the fundamental theorems based on the amalgamation of abutting intervals have been given recently by W. H. Young, *Proc. Lond. Math. Soc.* Ser. 2, vol. I, p. 240, and by Schoenflies, *Göttinger Nachrichten*, 1903. The proof given in the text, in § 73, is based upon the latter proof.

In any arbitrarily chosen interval a point P can be found which does not belong to a given non-dense closed set G, and an interval pq containing P in its interior can be found such that no internal point of pq belongs to G. For if no such interval could be found, P would be a limiting point of G, which is impossible. Let pq be the greatest interval containing P for which this holds, then p, q are limiting points of G, and therefore belong to G.

We now proceed to construct intervals such as pq in the remaining parts of the domain. Then, when every possible such interval has been constructed, there are no intervals complementary to them; and all the end-points of the intervals, together with the limiting points of such end-points, are the only points which are not internal to the intervals. These points are consequently the points of G.

The points of G consist in general of three classes:

(1) those which are common end-points of two intervals abutting on one another;

(2) semi-external points (see § 66), which are end-points of one interval and also limiting points on one side, of end-points; and

(3) external points, viz. such as are not end-points of intervals but are limiting points, on both sides, of end-points.

An end-point of the domain of the set may be regarded as belonging to (1) or (3) according as it is, or is not, an end-point of an interval.

Those points which belong to (1) are clearly isolated points of G. Hence if no such points exist, every point of G is a limiting point; and therefore G is perfect. The theorem has thus been proved that:—

Every non-dense perfect set G consists of the end-points of an everywhere-dense set of non-overlapping intervals no two of which abut on one another, together with the limiting points of these end-points.

The end-points of the domain are points of G, but not end-points of intervals.

If the set G is such that no semi-external points exist, then every interval abuts on another one at both its ends. In this case, all the points of G are either end-points of adjacent intervals, or limiting points, on both sides, of a sequence of such end-points, unless a or b be a limiting point, in which case it belongs to G. The end-points have the same cardinal number a as the rational numbers, since the set of intervals is enumerable. Moreover the external points form a finite set, or an enumerable set; because to each such external point there corresponds an enumerable set of end-points of which it is the limiting point, and in this correspondence any one end-point can correspond to at most two limiting points, one on each side of it. We thus have the theorem that:—

A non-dense closed set is enumerable if its complementary intervals are such that every one of them abuts on another one at each of its ends.

73. *Every non-dense closed set is, in general, made up of an enumerable set and of a perfect set.*

Let the intervals complementary to the set G be arranged in enumerable order, that of descending magnitude; we may denote them by $\delta_1, \delta_2, \dots \delta_n, \dots$. If G is not perfect, it contains isolated points, each of which is the common end-point of two adjacent intervals; let δ_{p_1} be the first of the intervals $\{\delta\}$ at an end of which there is such a point; let $\delta_{p'}$ be the interval which abuts on δ_{p_1} at that end. It may happen that the other end-point of $\delta_{p'}$ is also a common end-point of two intervals. If so, let $\delta_{p''}$ be the interval which abuts on $\delta_{p'}$, and so on: after a finite, or enumerable set, of such intervals

$$\delta_{p'}, \ \delta_{p''}, \ \delta_{p'''} \dots$$

we must arrive at an interval of which the end-point does not belong to G_ι, the set of isolated points of G, or else at an end-point of the domain of G; unless G is an enumerable set. It may happen that δ_{p_1} at its other end abuts on another interval; in that case we proceed, in the same manner as before, to find the intervals $\delta_{q'}, \delta_{q''}, \dots$ each of which abuts on another one. Now conceive all the intervals $\delta_{p'}, \delta_{p''}, \dots$, and if they exist, $\delta_{q'}, \delta_{q''}, \dots$ to be amalgamated with δ_{p_1} into one interval $\delta_{p_1}^{(1)}$, by removing all the common end-points. If any isolated points of G now remain, let δ_{p_2} be the first interval of $\{\delta\}$ after δ_{p_1}, of which an end-point is such a point; proceed as before, we then have an interval $\delta_{p_2}^{(1)}$ formed by amalgamating a finite or enumerable set of intervals. We proceed in this way, and thus form a set of intervals $\delta_{p_1}^{(1)}, \delta_{p_2}^{(1)}, \dots$ no end-points of which are points of G_ι.

Since $G = G_\iota + G^{(1)}$, where $G^{(1)}$ is the derivative of G, the set of intervals $\{\delta^{(1)}\}$ complementary to $G^{(1)}$ consists of the intervals $\delta_{p_1}^{(1)}, \delta_{p_2}^{(1)}, \dots$ and of any intervals of $\{\delta\}$ which remain after such intervals as $\delta_{p'}, \delta_{p''}, \dots \delta_{q'}, \delta_{q''}, \dots$ have been removed, and the $\delta_p^{(1)}$ substituted for the δ_p.

We proceed in a similar manner with $G^{(1)} \equiv G_\iota^{(1)} + G^{(2)}$, again removing a finite or enumerable number of the set $\{\delta^{(1)}\}$, and again with $G^{(2)}$, and so on. It may happen that the process comes to an end after a number n of such stages, either if $G_\iota^{(n)}$ does not exist, in which case $G^{(n)} = G^{(n+1)}$, and thus $G^{(n)}$ is perfect; or else, if $G^{(n)}$ does not exist, in which case, G being the sum of a finite number of enumerable sets $\Sigma G_\iota^{(r)}$, is itself enumerable. If the process does not come to an end for any finite value of n, we form the derivative $G^{(\omega)} = D(G^{(1)}, G^{(2)}, \dots G^{(n)}, \dots)$, which contains all the points common to all the derivatives of G of finite order. This set has been shewn in § 63, to exist, and to be a closed set; $G^{(\omega)}$ is then resolved as before into $G_\iota^{(\omega)} + G^{(\omega+1)}$, and we proceed further as before.

We obtain, by proceeding in this manner,

$$G = G_\iota + G_\iota^{(1)} + \dots + G_\iota^{(\omega)} + G_\iota^{(\omega+1)} + \dots + G_\iota^{(\beta)} + G^{(\beta+1)},$$

where β is a number of the first or second class. It will now be shewn that there must be some definite number β of the first or second class, for which this process comes to an end, either by $G_\iota^{(\beta)}$ containing no points, in which case $G^{(\beta)} = G^{(\beta+1)}$, so that $G^{(\beta)}$ is perfect; or else by $G^{(\beta+1)}$ containing no points, in which case G being the sum of an enumerable set of finite, or enumerable, sets, is itself enumerable. The $\{\delta\}$ contain all the indices 1, 2, 3, ... n, ...; from these indices we must remove a finite, or an enumerably infinite number, to obtain those indices which occur in the $\{\delta^{(1)}\}$; and again an enumerable set of indices must be removed from those which occur in the $\{\delta^{(n)}\}$, to obtain those which occur in the $\{\delta^{(\omega)}\}$. Now as the indices 1, 2, 3, ... n, ... are enumerable, the process of removing successively a finite, or enumerably infinite, set of them must cease for some order β of $\delta^{(\beta)}$, for otherwise a more than enumerable infinity of indices could be removed from the set 1, 2, 3, ... n, ... which is impossible; hence for some fixed number β of the second class all the indices must have been removed.

It has thus been shewn that, unless the given set G is enumerable, for some number β of the first or second class, $G^{(\beta)} = G^{(\beta+1)}$; and therefore $G^{(\beta)}$ is perfect. Thus G has been resolved into an enumerable set and a perfect one.

If for any value of β, $G^{(\beta)} \equiv 0$, the set G is enumerable.

74. The following theorem*, more general than that of § 73, includes the latter as a particular case. The proof here given may be taken as alternative to that of § 73.

If $P_1, P_2, \dots P_n, \dots P_\beta, \dots P_\alpha, \dots$ are all closed sets of points such that (1) if $\alpha_1 < \alpha_2$, all the points of P_{α_2} belong to P_{α_1}, and (2) if in any interval, any set P_α contains only a finite number of points, the set $P_{\alpha+1}$ contains no points in that interval; then either P_β must vanish for some definite number β of the first or second class, or else there is a definite number β such that P_β is a perfect set.

If for some number β, the set P_β vanishes, then P_γ vanishes for all values of γ which are $> \beta$.

Let us now suppose that there exists no number β such that P_β vanishes. In this case there exists a set of points which may be denoted by P_Ω, such that each point of the set belongs to P_β whatever number β may be. The set P_Ω is closed, for if p be a limiting point of the set, it is the limit of a sequence of points contained in P_β, whatever number β may be; hence p belongs to P_β, whatever β may be, and thus p itself belongs to P_Ω.

* See Baire, *Annali di Mat.*, Ser. 3, vol. III.

It will now be shewn that P_Ω contains no isolated points, and is therefore dense-in-itself. If P_Ω contains an isolated point p, a neighbourhood of p can be found which contains no point of P_Ω except p; let Q be that part of P_1 which is contained in this neighbourhood. In the neighbourhood considered, let us suppose a sequence of intervals $\delta_1, \delta_2, \ldots \delta_n, \ldots$ constructed, each one containing the next one and the point p, and such that δ_n converges to zero as n is indefinitely increased. Let $Q^{(n)}$ denote that part of Q which lies in δ_n but not in δ_{n+1}, then $Q = Q^{(1)} + Q^{(2)} + \ldots + Q^{(n)} + \ldots + p$. There must exist a number β_1, of the first or of the second class, for which $Q^{(1)}$ contains no point of P_{β_1}; otherwise Q_1 would contain points which belong to P_Ω, and this is not the case. Similarly, there exist numbers $\beta_2, \beta_3, \ldots \beta_n, \ldots$ such that $Q^{(2)}$ contains no points of P_{β_2}, and $Q^{(3)}$ contains no points of P_{β_3}, etc. Of the numbers $\beta_1, \beta_2, \ldots \beta_n, \ldots$, let γ_1 be the first which is $> \beta_1$, then let γ_2 be the first which is greater than γ_1, and so on; we have therefore a sequence $\gamma_1, \gamma_2, \ldots \gamma_n, \ldots$ of increasing numbers all of which belong to the set $\beta_1, \beta_2, \ldots \beta_n, \ldots$. This sequence $\gamma_1, \gamma_2, \ldots \gamma_n, \ldots$ is either finite, with say γ as the last, or else there is a limiting number γ of the second class which is greater than all of them, and therefore greater than all the numbers $\beta_1, \beta_2, \ldots \beta_n, \ldots$. The set Q can have no point except p which belongs to P_γ, hence since P_γ contains only one point in a certain interval, $P_{\gamma+1}$ contains no point in that interval, and does not contain p, which is contrary to the hypothesis.

It has now been shewn that P_Ω is closed and dense-in-itself; it is therefore perfect. Let us next consider the enumerable set of intervals which are complementary to P_Ω. For any one of these intervals there exists a number γ such that P_γ contains no point in the interior of the interval. As before, it is seen that there exists a number, of the first or the second class, which is greater than all these numbers γ; if this number be β, the set P_β contains no points which do not belong to P_Ω. It is thus seen that P_β is perfect, and

$$P_\beta = P_{\beta+1} = \ldots = P_\Omega.$$

The theorem has now been completely established.

75. *Every perfect set* has the cardinal number c of the continuum; and every closed infinite set has the cardinal number c, or else the cardinal number a of the rational numbers.*

Let the intervals whose internal points are the set $C(G)$, the complement of the perfect set G, be denoted by $\{\delta\}$; and let Δ denote the greatest, or one of the greatest in case of equality, of the intervals $\{\delta\}$. Let l, the whole interval (a, b) in which G lies, be divided into the three parts l_0, Δ, l_1 so that $l = l_0 + \Delta + l_1$, where l_0 is on the left, and l_1 on the right of Δ the greatest interval of $\{\delta\}$. Denote the greatest of the intervals $\{\delta\}$ in l_0, by Δ_0, and the

* Cantor, *Math. Annalen*, vol. XXIII.

greatest in l_1, by Δ_1; then the interval l_0 is divided by means of Δ_0 into three parts l_{00}, Δ_0, l_{01} in order from left to right, and the interval l_1 is divided by means of Δ_1 similarly into l_{10}, Δ_1, l_{11}. Proceeding in this manner to a further subdivision, let Δ_{pq} be the greatest of the intervals $\{\delta\}$ which lie in l_{pq}, where p, q each has one of the values 0 or 1; then l_{pq} is divided into three parts l_{pq0}, Δ_{pq}, l_{pq1}, and so on indefinitely. The intervals $\{\delta\}$ are thus arranged in the order $\Delta, \Delta_0, \Delta_1, \Delta_{00}, \Delta_{01}, \Delta_{10}, \Delta_{11}, \ldots$ and each interval of $\{\delta\}$ occurs at a definite place in the sequence. Consider a sequence of intervals

$$l,\ l_p,\ l_{pq},\ l_{pqr},\ \ldots,$$

where $p, q, r, \ldots$ all have definite values each of which is either 0 or 1. Each of these intervals is contained in the preceding one, and has one end-point in common with it; and the sequence determines a single point P which is interior to all the intervals of the sequence, unless, from and after some fixed index, all the indices are identical, in which case P is a common end-point of all the intervals after a fixed one. Hence since the point P is not interior to any of the intervals $\{\delta\}$, it is a point of G. Conversely, every point of G can be so determined by means of a sequence of intervals; for every point of G belongs either to l_0 or to l_1, and also to one of the four intervals l_{00}, l_{01}, l_{10}, l_{11}, and so on. The point P is the limiting point of the end-points of the intervals $\Delta_p, \Delta_{pq}, \Delta_{pqr}, \ldots$ with the indices the same as those of the sequence $l_p, l_{pq}, l_{pqr}, \ldots$ which determines the point.

Every number of the continuum (0, 1) is expressible in the dyad scale by means of a sequence $\cdot p, \cdot pq, \cdot pqr, \ldots$ where each of the numbers $p, q, r, \ldots$ is either 0 or 1; and all numbers are expressed uniquely in this manner, except those for which all the digits after some fixed one are 1, these numbers being also expressible by a sequence in which only 0 occurs after some fixed place. The numbers last mentioned correspond as indices of $l_p, l_{pq}, l_{pqr}, \ldots$ to a point of G which is an end-point of one of the intervals $\{\delta\}$; but in every other case a number in the dyad scale corresponds to a point of G which is not an end-point of the intervals $\{\delta\}$. Since the set of numbers of the continuum (0, 1) has the cardinal number c, it follows that the points of G form a set of the same cardinal number, because each point of G corresponds uniquely to a single number of the continuum, except that two points of G which are end-points of one interval correspond to a single number of the continuum. Every closed set which is not enumerable has been shewn to contain a perfect set as component; such a set has therefore the cardinal number c.

It will appear from the theory of order-types which will be discussed in the next Chapter, that the set of intervals $\{\delta\}$ which define a perfect set G, when taken in their order of position from left to right, have an order-type which is the same as η the order-type of the rational numbers which lie between 0 and 1, excluding 0 and 1 themselves, taken in their natural order in the continuum. It follows that a correspondence can be established between the intervals and the rational numbers, in which any two intervals correspond to two rational numbers that have the same order. If we take each rational number to correspond to the end-points of the corresponding interval, then each irrational number corresponds to a point of G which is a limiting point of end-points of intervals.

EXAMPLES.

1. Let x be a number given by $x = \frac{C_1}{3} + \frac{C_2}{3^2} + \frac{C_3}{3^3} + \dots + \frac{C_n}{3^n}$, where the numbers $C_1, C_2, \dots C_n$, have each one of the values 0, 2, and n has every integral value, and may also be indefinitely great. The set $\{x\}$ is a non-dense perfect set.

No number of the set lies between

$$\frac{C_1}{3} + \frac{C_2}{3^2} + \dots + \frac{0}{3^n} + \frac{2}{3^{n+1}} + \frac{2}{3^{n+2}} + \dots \quad \text{or} \quad \frac{C_1}{3} + \frac{C_2}{3^2} + \dots + \frac{1}{3^n},$$

and

$$\frac{C_1}{3} + \frac{C_2}{3^2} + \dots + \frac{2}{3^n};$$

these two numbers determine a complementary interval of the set, the interval being of length $\frac{1}{3^n}$. The number of complementary intervals of length $\frac{1}{3^n}$ is 2^{n-1}, hence the sum of all the complementary intervals is $\sum_{n=1}^{\infty} \frac{2^{n-1}}{3^n}$, which is unity. It is clear that the set of complementary intervals is everywhere-dense, and thus the set of points is non-dense. This example was constructed* by Cantor, and is the first example of a perfect non-dense set which has been purposely constructed.

2. Let us suppose that the numbers of the interval (0, 1) are expressed in the dyad scale, in the form $\cdot a_1 a_2 a_3 \dots a_n \dots$; where each a is either 0 or 1. Each number for which the a's all vanish, after some fixed one a_n, which must be 1, is also representable as an unending radix fraction, in which a_n is 0, and all the subsequent digits are 1. Let the numbers now be interpreted as if they were in the decimal scale. To each irrational number in the dyad scale, there corresponds a single number in the decimal scale, represented by the same digits. Of each rational number, there is a double representation in the dyad scale, and there correspond two numbers in the decimal scale, which define a complementary interval of the set of points which represents the numbers in the decimal scale. A perfect non-dense set of points is thus defined.

3. Taking a positive integer $m\,(>2)$, let the interval (0, 1) be divided into m equal parts, and exempt the last part from further subdivision. Divide each of the remaining $m-1$ intervals into m equal parts, and in each case exempt the last part from further subdivision. Let this operation be continued indefinitely. The points of division form a non-dense set; for if an interval d be taken anywhere in the interval (0, 1), k may be so chosen that $\frac{1}{m^k} < \frac{d}{2}$, and a segment $\left(\frac{a}{m^k}, \frac{a+1}{m^k}\right)$ entirely within d, can be determined. This

* See *Math. Annalen*, vol. XXI, p. 590.

segment is either an exempted interval, or its mth part is one. The end-points of the intervals, together with their limiting points, form a non-dense closed set, of cardinal number c.

4. As in* Ex. 3, let the interval (0, 1) be divided into m equal parts, and the last be exempted from further division. Then let the remaining $m-1$ parts each be divided into m^2 equal parts, the last of each being exempted from further division. Let the remaining parts be then divided into m^3 equal parts, the last of these in each case being exempted from further division. If this process be carried on indefinitely, the end-points of the divisions together with their limiting points, form a non-dense closed set, of cardinal number c.

5. Let $k_1, k_2, \ldots k_n, \ldots$ be a sequence of positive integers each of which is greater than unity, and defined according to any law.

It can be shewn† that every irrational number x, in (0, 1) can be uniquely represented in the form

$$x=\frac{c_1}{k_1}+\frac{c_2}{k_1k_2}+\ldots+\frac{c_n}{k_1k_2\ldots k_n}+\ldots$$

where $c_n<k_n$, and not all of the numbers $c_n, c_{n+1}, \ldots$ are zero, for any value of n.

It can further be shewn that

$$x=1-\frac{\eta_1}{k_1}-\frac{\eta_2}{k_1k_2}-\ldots-\frac{\eta_n}{k_1k_2\ldots k_n}-\ldots,$$

where $\eta_n=k_n-1-c_n$. If, from and after a certain value of n, the condition $c_n=k_n-1$, is always satisfied, then all the η_n vanish, and x is rational. It thus appears that the rational numbers are capable of a double representation in the form

$$x=\frac{c_1}{k_1}+\frac{c_2}{k_1k_2}+\ldots+\frac{c_n}{k_1k_2\ldots k_n}+\ldots;$$

(1) by the vanishing of all the c, after some fixed one, and (2) by the condition $c_n=k_n-1$ being satisfied from and after some fixed value of n.

If we now take those values of x, for which every c does not exceed some fixed integer λ, these values of x form a non-dense perfect set G_λ. It is easily seen that the interval of which the end-points are

$$\frac{c_1}{k_1}+\frac{c_2}{k_1k_2}+\ldots+\frac{c_n}{k_1k_2\ldots k_n}+\frac{\lambda}{k_1k_2\ldots k_{n+1}}+\frac{\lambda}{k_1k_2\ldots k_{n+2}}+\ldots$$

and

$$\frac{c_1}{k_1}+\frac{c_2}{k_1k_2}+\ldots+\frac{c_n+1}{k_1k_2\ldots k_n}$$

contains no points of the set in its interior, although these points belong to the set.

A particular case of this set consists of the numbers given by

$$x=\frac{c_1}{10}+\frac{c_2}{10^{1\cdot 2}}+\frac{c_3}{10^{1\cdot 2\cdot 3}}+\ldots+\frac{c_n}{10^{n!}}+\ldots,$$

where every c is $\leqq 9$. This set consists of the transcendental numbers first defined by Liouville‡.

* See H. J. S. Smith, *Proc. Lond. Math. Soc.* vol. VI, 1870.

† Brodén, *Math. Ann.* vol. LI.

‡ *Liouville's Journal*, vol. XVI, p. 133.

PROPERTIES OF THE DERIVATIVES OF SETS.

76. If a set is dense in any sub-interval of the domain in which it is contained, its derivative G' contains every point of the sub-interval, and is identical, so far as such sub-interval is concerned, with the totality of the points of the sub-interval; we confine ourselves therefore to the case in which G is a non-dense set, and consequently its derivatives are also non-dense.

The derivatives of transfinite orders have been defined in § 63; and it was there shewn that there is either a first derivative whose order is some number of the first class, or non-limiting number of the second class, or else that derivatives of all such orders exist, and have a set of points $G^{(\Omega)}$ in common.

It was shewn in § 73, that $G^{(1)}$ being a non-dense closed set, two cases arise:—

(1) If $G^{(1)}$ is enumerable, in which case G is also enumerable, then $G^{(\beta)}$ vanishes for some number β of the first or the second class. A set of this kind is called a *reducible* set.

(2) If $G^{(1)}$ is not enumerable, then there exists some number β, of the first or second class, for which $G^{(\beta)}$ is a perfect set, and is consequently identical with $G^{(\beta+1)}$, and with $G^{(\Omega)}$ as defined in § 63. The set $G^{(1)}$ is the sum of an enumerable set and the perfect set $G^{(\beta)}$. A set G which has this property, is said to be *irreducible*.

It should be observed that when $G^{(1)}$ is unenumerable, and consequently of cardinal number c, the same as the cardinal number of its perfect component, we are unable to make any inference as to the cardinal number of G itself. This may be a or c, or other cardinal number between the two, in case such a number exists.

THE CONTENT AND THE MEASURE OF SETS OF POINTS.

77. The theory of the content of a set of points in a finite linear domain was originated by Hankel*, and further developed by Harnack, Stolz, and by Cantor†, who extended the conception to the case of sets of points in a domain of any number of dimensions.

Suppose we have a linear set of points G in the finite interval (a, b), and conceive the interval to be divided into any finite number n_1 of sub-intervals, the greatest of which is Δ_1; let the sum of those ν_1 sub-intervals which have points of G either as interior points, or as end-points, be denoted by

* See Hankel, *Math. Annalen*, vol. XX; Stolz, *Math. Annalen*, vol. XXIII; Harnack, *Math. Annalen*, vol. XXV; Pasch, *Math. Annalen*, vol. XXX.

† *Math. Annalen*, vol. XXIII.

$S_{n_1\nu_1}$, where $\nu_1 \leqq n_1$; then $S_{n_1\nu_1} \leqq b-a$. Now suppose each sub-interval to be again divided into any number of parts, so that the whole interval (a, b) is now divided into n_2 sub-intervals $(n_2 > n_1)$ of which the greatest is Δ_2; and of these suppose $\nu_2 (\leqq n_2)$ to contain points of G as interior points or at their ends. Let $S_{n_2\nu_2}$ denote the sum of the ν_2 intervals; thus $S_{n_2\nu_2} \leqq S_{n_1\nu_1} \leqq b-a$.

Proceed in this manner to make further sub-divisions, so that at any stage there are n_r sub-intervals of (a, b), of which the greatest is of length Δ_r, and such that $S_{n_r\nu_r}$ is the sum of those ν_r intervals $(\nu_r \leqq n_r)$ which contain points of G. Let the process of further sub-division be continued indefinitely in any prescribed manner which is subject to the condition that $\Delta_1, \Delta_2, \dots \Delta_r, \dots$ is a sequence which has the limit zero. Then the numbers

$$S_{n_1\nu_1} \geqq S_{n_2\nu_2} \geqq S_{n_3\nu_3} \dots \geqq S_{n_r\nu_r} \dots$$

have a definite limit Σ to which $S_{n_r\nu_r}$ is arbitrarily near, for a sufficiently great value of r; and Σ may be equal to $b-a$, or to zero, or to a number between 0 and $b-a$.

It will now be shewn that the number Σ is independent of the original mode of sub-division of the interval (a, b), and of the mode in which the further sub-division is carried on, the sole restriction on the mode of formation of successive sub-divisions being that Δ_r, the greatest sub-interval of the rth sub-division, must have the limit zero when r increases indefinitely. Considering two different processes of sub-division, suppose Σ, Σ' to be the limits, in the two cases, of the sums of those sub-intervals which contain points of G. Suppose the first system of sub-division so far advanced that $S_{n_r\nu_r} - \Sigma < \epsilon$, where ϵ is an arbitrarily chosen positive number; and let the second system of sub-division be so far advanced that $\Delta'_s < d$, where d is an arbitrarily chosen positive number. If we conceive the two sets of points of division to coexist, we then have a further sub-division of (a, b), which may be considered as a continuation of either of the sub-divisions corresponding to $S_{n_r\nu_r}$ or to $S'_{n'_s\nu'_s}$. Suppose z is the sum of those of the new sub-intervals which contain points of G, then $z \leqq S_{n_r\nu_r}$. Of the sub-intervals in $S'_{n'_s\nu'_s}$, there can be at most $n_r - 1$ which are not sub-intervals of z; hence $S'_{n'_s\nu'_s} < z + n_r d$, and thus

$$\Sigma' < S'_{n'_s\nu'_s} < S_{n_r\nu_r} + n_r d < \Sigma + \epsilon + n_r d.$$

Now ϵ is arbitrarily small, and d is also arbitrarily small and independent of n_r; thus $\Sigma' \leqq \Sigma$. Similarly it can be proved that $\Sigma \leqq \Sigma'$; and thus Σ and Σ' must be equal.

We have now established the following theorem:—

If G be any given set of points in the interval (a, b), there corresponds to G a definite number Σ, which is such that all the points of G can be included in *a definite number of intervals whose sum exceeds Σ by less than an arbitrarily chosen positive number ϵ, the number of the intervals depending on ϵ.*

The number Σ is called the *content* of the set G, and the content may have any value between the two numbers 0, $b - a$, both inclusive.

Those sets of points for which the content is zero are of special importance in the Theory of Functions. A set of zero content is said to be an *unextended*, or a *discrete*, or an *integrable* set of points.

78. *The content of a set of points is the same as the content of its derivative.*

Let Σ' be the content of G' the derivative of a set G; then the points of G' may be included in the interiors of a finite set of intervals whose sum is less than $\Sigma' + \delta$, where δ is an arbitrarily chosen positive number. There can only be a finite number of points of G which do not fall within the intervals that include the points of G' in their interiors; and this finite number of points may be included in intervals whose sum is arbitrarily small, say ϵ. All the points of G are now included in a finite number of intervals whose sum is less than $\Sigma' + \delta + \epsilon$; and a series of diminishing values may be assigned to δ and ϵ, each sequence having the limit zero; and therefore both $\Sigma' + \delta + \epsilon$ and $\Sigma' + \delta$ converge to the value Σ'; which proves the theorem.

It follows from this theorem, that the content of any set is the same as that of any of its successive derivatives. In the case of a set which is of the first species, one of the derivatives contains only a finite number of points, and consequently the set must be of zero content.

79. A definition of the content of a set of points has been given by Cantor* which, though differing in form from that of Hankel and Harnack, is in reality equivalent to it. Instead of enclosing the points of the set G in a finite number of intervals, Cantor encloses each point of G in an interval 2ρ of which the point is the middle point, the number ρ being the same for each point of the set, those parts of intervals 2ρ which do not lie within (a, b) being disregarded. We have in this manner obtained an infinite number of overlapping intervals which contain all the points of G, and, as is clear, all the points of G', which is a closed set. If we replace this set of intervals by the set of non-overlapping intervals with the same interior points, each interval of this latter set is $\geqq 2\rho$. The set, which is non-overlapping, and equivalent to the infinite set, is consequently a finite set, the sum of whose lengths may be denoted by $\Pi(\rho, G)$. When ρ is diminished indefinitely, the number $\Pi(\rho, G)$, which cannot increase as ρ is diminished, must have a definite lower limit, which defines the content of either of the

* *Math. Annalen*, vol. XXIII.

sets G and G'. Since the infinite set of intervals which has been employed only covers a finite number of detached lengths, this definition is equivalent to that of Hankel and Harnack. Cantor applies this definition to the case of a set of points in a p-dimensional continuum, by enclosing each point in a "sphere" of radius ρ with its centre at the point; the content is then the lower limit of the volume of the continuum contained within the spheres.

The essential point in the above definition of the content of a set of points is that all the points are enclosed in a finite number of intervals which therefore enclose all the limiting points*; and the lower limit of the sum of these intervals is taken as defining the content of the set. If the points are enclosed from the commencement in an infinite number of intervals which are of unequal length, in accordance with some prescribed law, and the lengths of these intervals are then diminished, each one in a prescribed manner tending to the limit zero, then the limit of the sum of those parts of the interval which are included in the infinite set of intervals is not necessarily equal to the content as above defined. For example, let us consider the set of rational points in the interval (0, 1). These points can be arranged in enumerable order $P_1, P_2, P_3, \ldots$: now enclose P_1 in an interval of length $\frac{1}{2}\epsilon$, P_2 in an interval $\frac{1}{2^2}\epsilon$, &c., P_n in an interval of length $\frac{1}{2^n}\epsilon$, and so on; the total length covered by these intervals cannot exceed $\epsilon\Sigma\frac{1}{2^n}$ or ϵ, and this has the limit zero, as ϵ is diminished towards zero. On the other hand, the content of the set of rational points is the same as that of the derived set; but this consists of all the points of the interval (0, 1), and is therefore unity. In general, any enumerable set of points can be enclosed in an infinite number of intervals, which covers a length that is arbitrarily small, and has the limit zero; whereas the content of the set is not in general zero.

80. A completely satisfactory definition of the content of a set of points of the most general character should satisfy the condition of affording a consistent generalization of the notion of the length of a continuous linear set of points, or of the notions of area and volume, in the case of sets of points in two or three dimensions. In the case of closed sets, the definition given above leaves nothing to be desired in this respect; but in the case of open sets, the definition leads to consequences which are at variance with the fundamental properties of lengths, areas, and volumes, as understood for the case of continuous domains. If G_1, G_2 are two complementary sets of points in the continuous interval (0, 1), then, in order that the contents of the sets G_1, G_2 may accord with a generalization of the notion of length, their sum should be unity; however, when G_1 and G_2 are unclosed, this condition is in

* See Harnack, *Math. Annalen*, vol. XXIII, p. 241.

general not satisfied by the definition given above. For example, if G_1 consists of the rational points, and G_2 of the irrational points, each of the two sets G_1, G_2 has its content unity, the same as that of the continuum (0, 1) itself. Again, let us consider an everywhere-dense set of non-overlapping intervals contained in (0, 1); then the internal points of these intervals form an open set G_1, of which the derivative consists of all the points of the continuum (0, 1); the external and the end-points of the intervals forming a non-dense closed set G_2. It will be shewn subsequently that the everywhere-dense set of non-overlapping intervals can be so chosen that the limit of the sum of their lengths is an arbitrary number l, where l is subject to the condition $0 < l \leqq 1$; whereas the content of the set G_1 is, in accordance with the definition given above, always unity, and therefore may differ from the sum of the contents of the sets of points contained in the separate intervals. To obtain the content of the closed set G_2, cut off, from each of the intervals which define G_1, the $\frac{1}{2n}$th part of its length at each end; the limiting sum of the intervals so restricted is $l\left(1-\frac{1}{n}\right)$. Of these restricted intervals, a finite number can be so taken that their sum is $> l\left(1-\frac{1}{n}\right)-\epsilon$, and $< l\left(1-\frac{1}{n}\right)$, where ϵ is an arbitrarily chosen positive number. All the points of G_2 are now enclosed in the finite set of intervals which is complementary to the finite set of restricted intervals. The sum of these complementary intervals is $< 1 - l\left(1-\frac{1}{n}\right)+\epsilon$ and $> 1 - l\left(1-\frac{1}{n}\right)$; the sum has for its lower limit the number $1 - l$, which is therefore the content of G_2. The sum of the contents of G_1, G_2 is therefore not equal to unity, which is nevertheless the content of

$$G_1 + G_2 \equiv (0,\ 1).$$

81. For the reasons which have been explained and illustrated, the definition of the content of a set of points in the form given either by Hankel or by Cantor is appropriate only in the case of closed sets.

A theory has been developed by Borel* and Lebesgue†, and also by W. H. Young‡, of which the aim is to attach to each set of points a definite number called its *measure*, which shall be such as to form a natural extension of the notion of the length of a continuous interval, or of the notions of area and volume in higher dimensions. In this theory, certain postulates are made, by

* See his *Leçons sur la théorie des fonctions*.

† See the memoir "Integrale, Longeur, Aire" in the *Ann. di Mat.* Ser. III, vol. VII (1902).

‡ "Open sets and the theory of content," *Lond. Math. Soc. Proc.* Ser. II, vol. II, where a similar theory has been developed independently of the work of Lebesgue. Here the term "content" is used for Lebesgue's "measure"; the latter term has been adopted in the text in order to avoid confusion with the term "content" as used by Hankel, Harnack, and Cantor.

means of which definite measures are assigned to successive classes of sets of points. The question whether every set of points, however defined, has a measure, is left open, but it appears that all sets of points which have in point of fact been defined have definite measures; such sets are said to be *measurable*.

The problem of assigning definite measures to sets of points is taken to require that the measure of a set shall satisfy the following conditions:—

(1) Sets containing an infinite number of points exist of which the measure is not zero.

(2) Two congruent sets have the same measure.

Two sets are said to be *congruent* when, corresponding to every pair P, Q of points of the first set, there exist corresponding points P', Q' of the second set, such that the distance of P from Q is the same as that of P' from Q'. The second set is therefore the first in a displaced position.

(3) The measure of the sum of a finite number of sets, or the limiting sum of an enumerably infinite number of sets of points, is the sum, or the limit of the sum, of the measures of the different sets, provided that no two of the given sets have a point in common.

If G is any given set of points, let the points of G be enclosed in a finite or an infinite number of intervals; it has been shewn that these intervals are equivalent to an enumerably infinite, or a finite set of intervals which do not overlap. The sum, or the limit of the sum, of these non-overlapping intervals has a positive value depending upon the mode in which the intervals are constructed. This sum has a lower limit, when every possible choice of the intervals which enclose the given set is taken account of; and this lower limit $m_e(G)$ is defined to be the *exterior measure of the given set* G. Every set of points G has an exterior measure; and the points of the set can always be enclosed in an enumerably infinite, or finite, number of intervals, whose sum does not exceed $m_e(G) + \epsilon$, where ϵ is an arbitrarily chosen positive number.

Let $C(G)$ denote the set which is complementary to G relatively to the interval (a, b) of length l, in which G is contained; if $m_e\{C(G)\}$ denotes the exterior measure of $C(G)$, then $l - m_e\{C(G)\}$, which may also be denoted by $m_i G$, is defined to be the *interior measure of* G.

An equivalent definition of the interior measure of a set is the following*:—

The interior measure of a set is the upper limit of the content of its closed components.

For if the complementary set $C(G)$ be enclosed in a set of non-overlapping intervals whose sum, or limiting sum, is $m_e\{C(G)\} + \epsilon$, then the

* See W. H. Young, *loc. cit.*, p. 28.

closed set formed by the end-points of these intervals and the exterior points is a component of G, and has a content $l - m_e\{C(G)\} - \epsilon$. Since ϵ is arbitrarily small, this content is less than $l - m_e\{C(G)\}$ by less than any arbitrarily chosen small number; and thus $l - m_e\{C(G)\}$ is its upper limit.

Every set of points G has both an exterior and an interior measure. When the two are equal, the set G is said to be measurable, and the number

$$m_e(G) = m_i(G)$$

is defined to be the measure of G. The measure of a measurable set G may be denoted by $m(G)$.

It will be shewn that this definition satisfies the conditions which have been stated above, whenever it is applicable; and it will be shewn that the sets which ordinarily arise are measurable in accordance with this definition. Whether sets exist for which the external and internal measures are unequal, and, if so, whether it is possible to give a definition of the measure of such a set, so as to still satisfy the conditions given above, are questions which will not be here discussed.

A single point is clearly measurable, and has a zero measure. The points in a single continuous interval, or in a finite number of such intervals, are at once seen to be measurable, whether the intervals are open or closed; and the measure is the length of the interval, or the sum of the lengths of the intervals.

The condition that any set of points G is measurable, in the sense defined above, may be stated as follows:—

A set of points is measurable if its points can be enclosed in a finite, or enumerably infinite, number of non-overlapping intervals α, and the complementary set can similarly be enclosed in intervals β, such that the sum, or limiting sum, of the common parts of α and β is arbitrarily small.

Any enumerable set of points is measurable, and has zero for its measure.

For if $P_1, P_2, \ldots P_n, \ldots$ are the points, they may be enclosed in intervals of lengths $\frac{1}{2}\epsilon, \frac{1}{2^2}\epsilon, \ldots \frac{1}{2^n}\epsilon, \ldots$, and the sum of the lengths of the equivalent non-overlapping intervals is $\leq \epsilon$, which is arbitrarily small. The exterior measure being zero, the set is measurable.

The points contained in the interior of an enumerably infinite set of non-overlapping intervals form a measurable set, whose measure is the limiting sum of the intervals.

This theorem has been shewn above to be not in general true of the content of such a set, as the content is defined in § 77.

If λ is the length of one of the intervals, cut off $\frac{1}{2n}$ of λ from each end;

we have then a curtailed interval of length $\lambda\left(1-\frac{1}{n}\right)$; do the same, taking the same value of n, with each interval of the set. Then, if we remove all these curtailed intervals from (a, b), we have all the points complementary to the given set enclosed in a set of intervals, such that the part which is common to them and to the given set of intervals is at most $\frac{1}{n}\Sigma\lambda$, which may be made as small as we please by taking n large enough. The condition that the given set of points is measurable is therefore satisfied; and it follows that the closed set which is complementary to the given set is also measurable.

It will hereafter be proved that every closed set is definable as the complementary set of the points interior to an enumerable set of intervals. From this it follows that every closed set is measurable, and that its measure is identical with the content as defined in § 77. The measure of the set of points interior to the intervals is however not in general identical with its content, as we have shewn in § 80.

82. It will now be proved that *if* $G_1, G_2, \ldots G_n, \ldots$ *are a finite, or enumerably infinite, number of sets which are measurable, then* $M(G_1, G_2, \ldots G_n, \ldots)$ *is measurable; and, if no point belongs to more than one of the sets, the measure of* $G_1 + G_2 + \ldots + G_n + \ldots$ *is the sum of the measures of* $G_1, G_2, \ldots G_n, \ldots$; *the definition of the measure thus satisfying the condition* (3).

Let G_1 and its complement $C(G_1)$ be enclosed in sets of intervals α_1, and β_1, each of which sets is non-overlapping, and such that the total length of the parts common to α_1, β_1 is the arbitrarily small number ϵ_1. Let G_2 and $C(G_2)$ be similarly enclosed in sets of intervals α_2, β_2 which have a common part ϵ_2, arbitrarily small; and let α_2', β_2' be the parts of α_2, β_2 which they have in common with β_1. For G_3 and $C(G_3)$, we similarly take α_3, β_3, which have a common part ϵ_3, arbitrarily small; α_3', β_3' are the parts which α_3, β_3 have in common with β_2': and so on.

The points of $M(G_1, G_2, G_3, \ldots)$ can be enclosed in the intervals

$$\alpha_1, \alpha_2', \alpha_3', \ldots;$$

moreover $C\{M(G_1, G_2, G_3, \ldots)\}$ has all its points enclosed in β_ι', whatever value ι may have; therefore the two sets of intervals have the common part

$$\epsilon_1 + \epsilon_2 + \ldots + \epsilon_\iota + m(\alpha'_{\iota+1}) + m(\alpha'_{\iota+2}) + \ldots.$$

The series $\Sigma m(\alpha')$ being convergent, since each term is positive, and the sum has a finite upper limit, ι can be chosen such that $m(\alpha'_{\iota+1}) + \ldots$ is less than ϵ, where ϵ is an arbitrarily small number; and $\epsilon_1, \epsilon_2, \ldots$ may be chosen so that $\epsilon_1 + \epsilon_2 + \ldots < \epsilon$: therefore the common part of the two sets of intervals is $< 2\epsilon$, and thus $M(G_1, G_2, G_3, \ldots)$ is measurable.

If G_1, G_2, G_3, ... have no points in common, we see that

$m(G_1+G_2+G_3+\ldots)$ differs from $m(\alpha_1)+m(\alpha_2')+m(\alpha_3')+\ldots$

by less than $\epsilon_1+\epsilon_2+\epsilon_3+\ldots$; and therefore

$$m(G_1+G_2+G_3+\ldots)=m(G_1)+m(G_2)+\ldots,$$

and thus the measure of the sum of a finite, or enumerably infinite, number of measurable sets satisfies the condition (3).

If one set G_1, which is measurable, contains another set G_2 which has the same property, then G_1-G_2 is measurable.

The complement of G_1-G_2 consists of G_2 together with $C(G_1)$, hence G_1-G_2 is measurable. Further, since

$G_1=(G_1-G_2)+G_2$, we have $m(G_1-G_2)=m(G_1)-m(G_2)$.

If G_1, G_2, G_3, ... are a finite, or enumerably infinite, number of measurable sets, then the set $D(G_1, G_2, G_3, \ldots)$ of points common to all of them is measurable.

For the complement of $D(G_1, G_2, \ldots)$ is $M\{C(G_1), C(G_2), \ldots\}$, and since $C(G_1)$, $C(G_2)$, ... are all measurable, it follows that the complement of $D(G_1, G_2, \ldots)$ is measurable, and therefore that the set itself is measurable.

If G_1, G_2, G_3, ... G_n, ... are an enumerably infinite number of measurable sets, and H is the set of points each of which belongs to an infinite number of the given sets, then H is measurable.*

For the set $C(H)$ complementary to H, consists of those points which belong to none, or only to a finite number, of the sets G_1, G_2, ... G_n, ...; and hence $C(H)$ consists of the points which belong to one or more of the sets L_1, L_2, ... L_n, ..., where L_n denotes the set $D\{C(G_n), C(G_{n+1}), \ldots\}$ which consists of the points common to all the sets $C(G_n)$, $C(G_{n+1})$, ... The sets L_n are all measurable, hence $C(H)$ is measurable; and therefore H is measurable.

If G_1, G_2, G_3, ... G_n, ... are an enumerably infinite number of measurable sets, and K is the set of points each of which belongs to all the sets G_n, G_{n+1}, ..., where n has a definite value for each point of the set K; then the set K is measurable.*

For the set $C(K)$, which is complementary to K, is the set of points each of which belongs to an infinite number of the measurable sets $C(G_1)$, $C(G_2)$, ... $C(G_n)$, ...; and hence, by the last theorem, $C(K)$ is measurable. Therefore K is a measurable set.

* See Borel's *Leçons sur les fonctions de variables réelles*, p. 18. The set H is named, by Borel, the "ensemble limite complet," and the set K the "ensemble limite restreint," of the given sequence of sets.

83. All the sets which we have so far proved to be measurable were obtained from two fundamental sets, the single point, and the single interval, open or closed, by taking a finite, or enumerably infinite, number of these fundamental sets, and by taking the set common to a finite, or enumerably infinite, number of the sets so obtained, or by taking the complements of the measurable sets so obtained.

It can be shewn that measurable sets may exist which are not definable in the manner we have described. It will be subsequently shewn that perfect sets exist whose measure is zero; any component whatever of such a set has its external measure zero, and is therefore measurable. The cardinal number of all such components is usually regarded as greater than the cardinal number of the continuum, which is the cardinal number of all the sets obtainable in the manner indicated above. It follows that other measurable sets may be definable besides those obtained by the processes we have described. Whether every set which can be defined is measurable, is a point which has not been settled.

If G be any measurable set whatever, the points of G can be enclosed in a set of intervals α_ι, whose sum is $m(G) + \epsilon_\iota$; where ϵ_ι is one of a sequence

$$\epsilon_1, \epsilon_2, \ldots \epsilon_\iota, \ldots$$

of positive decreasing numbers which converge to zero. The set G_1 of the points which are common to all the sets of intervals $\alpha_1, \alpha_2, \ldots \alpha_\iota, \ldots$ is measurable, and its measure is $m(G)$; also the set G_1 contains G as a component. The set $G_1 - G$ has measure zero, and its points can be enclosed in a set of intervals β_ι contained in α_ι, and of measure ϵ_ι. The set H of points common to all the β_ι is measurable, and of measure zero; and the set $G_2 = G_1 - H$ is measurable, and has $m(G)$ for measure. It has thus been shewn that every measurable set G is contained in another measurable set G_1, and also contains a third measurable set G_2; where G_1, G_2 are measurable sets of the type definable as the set common to an enumerable number of sets of intervals. The measures of G_1, G_2 are both $m(G)$.

84. A definition has been employed by Jordan*, and by Peano†, of the measure of a set of points, which differs from the one which has been developed above. It is applicable to sets of points in space of any number of dimensions.

Let G be a set of points in a domain E, and let $C(G)$ denote the set complementary to G; then a point of G which is not a limiting point of $C(G)$ is said to be an *interior point of* G; and a point of $C(G)$ which is not a limiting point of G is said to be an *interior point of* $C(G)$.

* *Liouville* (4), vol. VIII (1892); also *Cours d'Analyse*, vol. I, p. 28.
† *Applicazioni geom. del. calc. infinit.* (1887), p. 153.

Every point of E, which is an interior point neither of G nor of $C(G)$, is said to be a *point of the frontier of* G. Every such point is either a point of G, which is at the same time a limiting point of $C(G)$, or else it is a point of $C(G)$, which is also a limiting point of G; and the aggregate of all such points constitutes the frontier of G.

Divide E into any finite number of continuous parts, consisting of linear intervals or of rectangular cells, some of which may if necessary extend beyond the domain E; and let Σ_1 be the sum of those parts which are such that every point of each of them is an interior point of G. When the number of the continuous parts is increased indefinitely, in such a manner that the greatest of them converges to the limit zero, it can be shewn that Σ_1 converges to a definite limit S_1. If the sum Σ_2 of those parts of E is taken, each of which contains at least one interior point of G, or a point on the frontier of G, it can be shewn that Σ_2 converges to a fixed number S_2.

The number S_1 is called the *interior extent* of G, and the number S_2 is called the *exterior extent* of G; when S_2 is equal to S_1, the set G is said to be measurable, and $S_1 \equiv S_2$ is its measure. The exterior extent of a set G is identical with its content. In accordance with this definition, a set which does not contain any part which is a continuum has its interior extent zero; and such a set is only measurable when its content is also zero.

It can be shewn that, in those cases in which a set is measurable, both in accordance with this definition and with that of Borel, the measure is the same in the two cases. A set which is measurable in accordance with Jordan's definition is also measurable in accordance with that of Borel; but the converse does not hold. The definition of Borel will accordingly be employed, as being the more widely applicable of the two definitions.

THE CONTENT OF CLOSED SETS.

85. *The content of a non-dense closed set is zero, in case the set is enumerable; and in case the set is unenumerable, its content may be zero, or may have any value less than the length of the whole interval in which the set is contained.*

The content of a closed set being the same as its measure, its content is the sum of the content of its perfect component and the measure of its enumerable component; and this last is zero. If the set is enumerable, it has no perfect component, and therefore its content is zero. Since the content of any closed set is the same as that of its perfect component, it will suffice to consider the content of a non-dense perfect set.

Let l be the whole length of the interval in which the set is contained, the end-points of which interval belong to the set; let $\delta_1, \delta_2, \delta_3, \ldots$ be the intervals complementary to the set, arranged in descending order of magnitude. Of these intervals no two abut, and the set of intervals is everywhere-dense.

Let $\delta_1 = \lambda_1 l,\ \delta_2 = \lambda_2(l - \delta_1),\ \delta_3 = \lambda_3(l - \delta_1 - \delta_2), \dots$

$$\delta_n = \lambda_n(l - \delta_1 - \delta_2 - \dots - \delta_{n-1}) \dots;$$

thus $\lambda_1, \lambda_2, \lambda_3, \dots$ are proper fractions.

We have $\delta_2 = \lambda_2(1 - \lambda_1)l,\ \delta_3 = \lambda_3(1 - \lambda_1)(1 - \lambda_2)l, \dots$

$$\delta_n = \lambda_n(1 - \lambda_1)(1 - \lambda_2)\dots(1 - \lambda_{n-1})l,$$

hence $l - (\delta_1 + \delta_2 + \dots + \delta_n) = (1 - \lambda_1)(1 - \lambda_2)\dots(1 - \lambda_n)l.$

The content of the set is therefore* l multiplied by the limit of the product

$$(1 - \lambda_1)(1 - \lambda_2)\dots(1 - \lambda_n).$$

The values of $\lambda_1, \lambda_2, \dots \lambda_n, \dots$ can be so chosen that the content is zero; for example, we may take $\lambda_1 = \lambda_2 = \dots = \lambda_n = \dots$. Those perfect sets, and the related closed ones, which have content zero, are of special importance in the Theory of Functions.

The values of $\lambda_1, \lambda_2, \dots$ may be so chosen that the content of the set is arbitrarily nearly equal to l. For example, let

$$\lambda_1 = \theta,\ \lambda_2 = \theta/2^2,\ \lambda_3 = \theta/3^2, \dots \lambda_n = \theta/n^2, \dots$$

where θ is a fixed positive fraction, then the content of the set is

$$l \sin(\pi\sqrt{\theta})/\pi\sqrt{\theta},$$

and this may be made as nearly equal to l as we please, by choosing a sufficiently small value of θ. That in the interval l we may place an indefinitely great number of non-abutting intervals whose sum is arbitrarily small, and so that no interval exists whose points are all external to the set of intervals, is one of the paradoxes of the subject.

86. A closed set of the most general type is obtained by adding to a non-dense closed set the internal points of some of the complementary intervals. For, if a closed set be everywhere-dense in any interval (α, β) contained in (a, b), the interval in which the set exists, it is clear that every point of (α, β) belongs to the closed set. If the interior points of (α, β) be removed from the set, the remaining set is still closed. We may conceive this process of removing the interior points of intervals in which the closed set is everywhere-dense, to be continued, until a closed set remains which is dense in no interval. It has been shewn in § 72, that every non-dense closed set is definable as the end-points of an everywhere-dense enumerable set of non-overlapping intervals, together with the limiting points of these end-points. It has thus been shown that *every closed set is definable as the complementary set of the points interior to a finite, or enumerable, set of non-overlapping intervals, not necessarily everywhere-dense.*

* Harnack, *Math. Annalen*, vol. XIX.

It follows from this result, that the content of a closed set which is dense in some parts of the interval, is the sum of the content of the non-dense closed set from which it can be derived, and of the lengths of those intervals all the points of which belong to the closed set.

EXAMPLES.

1. The perfect set of points defined by $x = \frac{c_1}{3} + \frac{c_2}{3^2} + \ldots + \frac{c_n}{3^n} + \ldots$, where the numbers $c_1, c_2, \ldots$ have each one of the values 0, 2 (see Ex. 1, § 75), has the content zero. For the limit of the sum of the complementary intervals is unity.

2. The non-dense closed set considered in Ex. 3, § 75, has the content zero. For, after k operations, the sum of the exempted segments is

$$\frac{1}{m} + \frac{m-1}{m^2} + \frac{(m-1)^2}{m^3} + \ldots + \frac{(m-1)^{k-1}}{m^k}, \quad \text{or} \quad 1 - \left(\frac{m-1}{m}\right)^k.$$

When k is increased indefinitely, the limit of the sum of the free intervals is 1.

3. The non-dense closed set considered in Ex. 4, § 75, has a content between 0 and 1. After k operations, the sum of the exempted segments is

$$\frac{1}{m} + \frac{m-1}{m^3} + \frac{(m-1)(m^2-1)}{m^6} + \ldots + \frac{(m-1)(m^2-1)\ldots(m^{k-1}-1)}{m^{\frac{1}{2}k(k+1)}}$$

or

$$1 - \left(1 - \frac{1}{m}\right)\left(1 - \frac{1}{m^2}\right)\ldots\left(1 - \frac{1}{m^k}\right).$$

The limit of the sum of the exempted intervals is $1 - \prod_{k=1}^{\infty}\left(1 - \frac{1}{m^k}\right)$, and therefore the content of the set of points is $\prod_{k=1}^{\infty}\left(1 - \frac{1}{m^k}\right)$, which is between 0 and 1, depending upon the value of m. By taking m sufficiently great, the content of the set may be made arbitrarily near to unity.

ASCENDING SEQUENCES OF CLOSED SETS.

87. Let $G_1, G_2, \ldots G_n, \ldots$ denote a sequence of non-dense closed sets contained in the interval (a, b), such that each set contains the preceding one as a component; then the set G_ω, the limiting set of the sequence, is defined as a set such that any point P belonging to it is a point of some definite G_n, and consequently also of all the subsequent sets of the sequence; and, further, that every point which belongs to any G_n is a point of G_ω.

The limiting set G_ω is not necessarily a closed set; and it may, or may not be everywhere-dense in (a, b). Thus its derivative G_ω' may be a non-dense closed set, or it may be the continuum (a, b). To make it clear that G_ω may have limiting points which do not belong to any G_n, we observe that, if $P_1Q_1, P_2Q_2, \ldots P_nQ_n, \ldots$ are complementary intervals of $G_1, G_2, \ldots G_n, \ldots$ respectively, such that each interval contains the next as an interior interval, the lengths of the intervals may converge to zero; in that case there exists a point p which is interior to all the above intervals, but does not belong to any of the sets $G_1, G_2, \ldots G_n, \ldots$. Since this point p is interior to a complementary interval of each set, it is the limiting point of each of the sequences $P_1, P_2, \ldots P_n, \ldots, Q_1, Q_2, \ldots Q_n, \ldots$ of points all of which belong to G_ω: it thus appears that p is a point of G_ω', but not of G_ω; or G_ω is an open set.

It may happen that the sequence of intervals $P_1Q_1, P_2Q_2, \ldots P_nQ_n, \ldots$ is such that the limit of P_nQ_n is not zero. In that case $P_1, P_2, \ldots P_n, \ldots$ may have a limiting point p different from any P_n; and $Q_1, Q_2, \ldots Q_n, \ldots$ a different limiting point q different from any Q_n. The points p and q are in this case not points of G_ω, but are points of G_ω'; and the open interval pq is a complementary interval of G_ω'.

If, from and after some value n_1 of n, all the points P_n are coincident with p, then p is a point of G_ω; and a similar remark applies to q.

To shew that the set G_ω may be everywhere-dense in (a, b), let G_1 be a perfect set; in each complementary interval of G_1 place a perfect set similar to G_1, *i.e.* identical with G_1, except that the distances of every pair of points are reduced in the ratio of the length of the complementary interval to that of (a, b): we have now a new perfect set G_2. Place in each complementary interval of G_2 a set similar, in the same sense, to G_1; we thus obtain G_3; and so on indefinitely. The resulting limiting set G_ω is everywhere-dense: for, if the greatest interval in G_1 is θ times the length of (a, b), then the greatest interval in G_n is θ^n times the length of (a, b); and this is arbitrarily small as n increases. Hence, in any interval whatever taken in (a, b), for a sufficiently great value of n, complementary intervals, and therefore points of G_n, are contained; and therefore G_ω is everywhere-dense.

In the case in which G_ω is closed, it will subsequently appear that it must be non-dense. In this case a complementary interval of G_ω is either the whole, or a part, of a complementary interval of G_n, whatever value n may have; for, if pq be such a complementary interval of G_ω, no interior point of pq belongs to any G_n. For some value n_1 of n, p belongs to G_{n_1} but not to G_{n_1-1}; and for some value n_2 of n, q belongs to G_{n_2}, but not to G_{n_2-1}; and therefore, if m is the greater of the numbers n_1 and n_2, pq is a complementary interval of G_n, for every n which is $\geqq m$.

When G_ω is open but non-dense, it can be shewn* that, *corresponding to any complementary interval pq of G_ω', a number m can be found such that, for every n which is $\geqq m$, G_n has a complementary interval $p'q'$ which contains pq, and exceeds it by less than an arbitrarily small number η.*

* W. H. Young, *Proc. Lond. Math. Soc.* vol. XXXV, p. 275.

Increase pq on each side by $\frac{1}{2}\eta$, and let PQ be the interval so increased. If p and q are not both points of G_ω, an interval $p'q'$ contained in PQ, and containing pq, can be found such that p', q' are both points of G_ω. Now p', q' are both points of some set G_m; then either $p'q'$ is a complementary interval of G_m, or else G_m has a complementary interval whose end-points lie in (p', p) and (q, q') respectively. In either case G_m has a complementary interval which exceeds pq by less than η, and contains pq.

If ϵ, σ are arbitrarily small positive numbers chosen independently of one another, an integer m can be found such that, provided $n \geqq m$, the difference between the sum $s_n(\epsilon)$ of those complementary intervals of G_n, each of which is $\geqq \epsilon$, and of those of G_ω' which are $\geqq \epsilon$, is $< \sigma$.

Let s be the number of such complementary intervals of G_ω'; then for each such interval a value of n can be found such that G_n has a complementary interval which contains that of G_ω', and exceeds it by less than σ/s. Hence a value ν can be found of n, such that G_ν has a number of complementary intervals all $\geqq \epsilon$, and whose sum exceeds the sum of those of G_ω' which are $\geqq \epsilon$, by less than σ.

It may however happen that G_ν has other complementary intervals which are $\geqq \epsilon$, but of course only a finite number of such intervals. Let PQ be such an interval of G_ν; then PQ contains no interval of G_ω' which is $\geqq \epsilon$. In PQ we can take a finite number of points of G_ω, say $p_1, p_2, \dots p_n$, such that $Pp_1, p_1p_2, p_2p_3, \dots$ are each less than ϵ. If we treat each of the finite number of intervals of G_ν, such as PQ, in a similar manner, there exists a value m of n $(m > \nu)$ such that all the points p for every interval PQ are points of G_m; then the set G_m has no complementary intervals which are $\geqq \epsilon$, except such as contain the intervals of G_ω' which are $\geqq \epsilon$; and this proves the theorem.

It is clear that, in the case when G_ω is closed, the above theorem reduces to the simpler form, that corresponding to an arbitrary ϵ, a number m can be found, such that, for $n \geqq m$, those intervals complementary to G_n, which are $\geqq \epsilon$, are identical with those of G_ω, which are $\geqq \epsilon$.

88. The set G_ω may be regarded as the sum of the sets G_1, $G_2 - G_1$, $G_3 - G_2, \dots$ each of which contains no points which belong to the preceding ones. Since G_1, G_2, $G_3, \dots$ are measurable sets, it follows from § 82, that $G_2 - G_1$, $G_3 - G_2, \dots$ are also measurable; and thus that G_ω is measurable, its measure being the limiting sum of the measures $m(G_1)$, $m(G_2 - G_1)$, $m(G_3 - G_2), \dots$. Now it has been proved that

$$m(G_n - G_{n-1}) = m(G_n) - m(G_{n-1}),$$

and therefore the limiting sum of the measures of the sets is the limit of $m(G_n)$, which is a number that does not decrease as n increases. It thus appears that $m(G)$ is the limit of $m(G_n)$ as n is indefinitely increased; and hence it has been proved that:—

The measure of the limit of a sequence of non-dense closed sets is the upper limit of the measures of the sets of the sequence.

The measure of a closed set being identical with its content, we obtain Osgood's theorem* that:—

If the limit of a sequence of closed sets is itself a closed set, then the content of the limiting set is the upper limit of the contents of the sets of the sequence.

It should be observed that, when G_ω is not closed, it is in general not true that the content of G_ω, or of G_ω', is the limit of the content of G_n. For example, if all the sets G_n have zero content, the points of each G_n can be enclosed in a *finite* number of intervals of arbitrarily small sum; but this is not in general true of G_ω, unless G_ω is closed. The points of G_ω can however in this case be enclosed in an indefinitely great number of intervals whose limiting sum is arbitrarily small.

The content of any closed component of a measurable set G_ω cannot exceed the measure of G_ω; we have therefore the theorem that:—

If G_ω is the limiting set of a sequence of non-dense closed sets $G_1, G_2, \ldots G_n, \ldots$ each one of which contains the preceding one, then no closed component of G_ω can have content greater than the limit of the content of G_n.

An important particular case of this theorem arises when all the sets G_n have zero content; in that case every closed component of G_ω has zero content.

If ϵ is an arbitrarily small number, and $s_n(\epsilon)$, $s(\epsilon)$ are the sums of those complementary intervals of G_n, G_ω', respectively, each of which is $\geqq \epsilon$; and $R_n(\epsilon)$, $R(\epsilon)$ the sums of those complementary intervals of G_n, G_ω' each of which is $< \epsilon$, we have

$$s_n(\epsilon) + R_n(\epsilon) = m\{C(G_n)\}, \quad s(\epsilon) + R(\epsilon) = m\{C(G_\omega')\};$$

hence $$m\{G_\omega' - G_n\} = \{s_n(\epsilon) - s(\epsilon)\} + \{R_n(\epsilon) - R(\epsilon)\}.$$

Now, as we have above shewn, if $n \geqq m$, where m depends on ϵ,

$$s_n(\epsilon) - s(\epsilon) < \sigma;$$

and we can choose ϵ so that $R(\epsilon)$ is as small as we please. Therefore we see that† the necessary and sufficient condition for the measure of G_ω' being the same as that of G_ω is that ϵ can be so chosen that, from and after some fixed value of n, $R_n(\epsilon)$ may be less than an assigned arbitrarily small number. If $R_n(\epsilon)$ has not the limit zero, when n is indefinitely increased, it is certain that G_ω is unclosed, and has a measure less than the content of G_ω'.

* *American Journal of Math.*, vol. XIX, p. 178.

† W. H. Young, *Proc. Lond. Math. Soc.* vol. XXXV, p. 284.

SETS OF THE FIRST AND OF THE SECOND CATEGORY.

89. *If $P_1, P_2, \dots P_n, \dots$ is a sequence of non-dense closed sets, the set $M(P_1, P_2, \dots P_n, \dots)$, which contains all points belonging to one at least of the sets, is said to be a set of the first category.*

A set of the first category can be exhibited as the limit of a sequence of non-dense closed sets each of which contains the preceding one. For such a sequence is

$$P_1,\ M(P_1, P_2),\ M(P_1, P_2, P_3), \dots M(P_1, P_2, \dots P_n, \dots), \dots.$$

It is clear that a set of the first category is of cardinal number a or c, the former in case all the sets P are enumerable, and the latter in case some or all are unenumerable.

It has been shewn in § 87 that a set of the first category may be everywhere-dense in its domain; or it may be non-dense.

A set which is complementary to a set of the first category is said to be of the second category.*

It will be shewn that such a set is not of the first category.

In the first place, the set complementary to a set of the first category is everywhere-dense. For if (α, β) is any interval of the domain, there exists in the interior of (α, β) an interval (α_1, β_1), which contains no points of P_1; and in (α_1, β_1) is contained an interval (α_2, β_2) which contains no points of $M(P_1, P_2)$; and so on: there is consequently a point interior to all the intervals (α, β), (α_1, β_1), $(\alpha_2, \beta_2), \dots (\alpha_n, \beta_n), \dots$ which is not a point of $M(P_1, P_2, \dots P_n, \dots)$; hence the complementary set is everywhere-dense. It follows from this result that the continuum (a, b) is not a set of the first category.

Next, suppose if possible that the set complementary to

$$M(P_1, P_2, \dots P_n, \dots)$$

is itself the limit of a sequence $Q_1, Q_2, \dots Q_n, \dots$ of non-dense closed sets. The sets $P_1 + Q_1, P_2 + Q_2, \dots P_n + Q_n, \dots$ are all closed non-dense sets, and their limiting sum, which is of the first category, is identical with the continuum; but this we have shewn to be impossible. Hence the complement of a set of the first category is not of the first category.

A set of the second category has the cardinal number c of the continuum.

This is obvious in case the limiting set G_ω of a sequence $G_1, G_2, \dots G_n, \dots$ of non-dense closed sets, each of which is contained in the following one, has any complementary intervals. We can therefore confine ourselves to the case in which G_ω is everywhere-dense; in which case the greatest intervals

* The distinction between sets of the first and second category is due to Baire, *Annali di Mat.* (3), vol. III, p. 65.

$\delta_1, \delta_2, \ldots \delta_n, \ldots$ which are complementary to $G_1, G_2, \ldots G_n, \ldots$ respectively, form a sequence which converges to zero. Taking any complementary interval Δ of G_1, a number n_1 can be found such that Δ contains at least two intervals complementary to G_{n_1} in its interior, and these we denote by Δ_0, Δ_1. Again $n_2 > n_1$, can be found such that in each of the intervals Δ_0, Δ_1 are contained at least two intervals complementary to G_{n_2}; those interior to Δ_0 we denote by Δ_{00}, Δ_{01}, and those interior to Δ_1 by Δ_{10}, Δ_{10}. Proceeding in this way we obtain a sequence of intervals

$$\Delta_p, \Delta_{pq}, \Delta_{pqr}, \ldots$$

each of which contains the next in its interior, and $p, q, r, \ldots$ have definite values each of which is either 0, or 1. The point interior to all the intervals of this sequence is a point of G_ω' which does not belong to G_ω, unless, from and after some fixed index, all the indices are alike 0, or alike 1. Therefore those points of G_ω' which do not belong to G_ω have a (1, 1) correspondence with all those numbers between 0 and 1, expressed in the dyad scale, which do not contain identical digits from and after some fixed one; and it thus appears that the points complementary to G_ω form a set of cardinal number c.

Any two sets of the second category have in common a set of points which is also of the second category.

If G_ω is the limit of G_n, and H_ω is the limit of H_n, where G_n, H_n are the nth sets of two ascending sequences of non-dense closed sets, then the set $M(G_n, H_n)$, which is also closed, has for its limit a set of the first category. But the complement of this set is the set common to the two sets of the second category which are complementary to G_ω, H_ω; and this common set is itself of the second category.

The definitions of sets of the first and of the second category can be extended to the case in which all the sets concerned are contained in a perfect set H, which takes the place of the continuous interval (a, b).

If $G_1, G_2, \ldots G_n, \ldots$ are closed sets all contained in H, and each one non-dense in H, then the set $M(G_1, G_2, \ldots G_n, \ldots)$ is said to be a *set of the first category relatively to the perfect set H*, and its complement relatively to H is said to be a *set of the second category relatively to H.*

The perfect set H may be non-dense in the continuum; or it may contain continuous intervals, finite or indefinitely great in number.

It will appear, from the theory of order-types developed in the next chapter, that the points of any perfect set can be made to correspond uniquely with the points of a continuous interval (a, b), in such a manner that the relative order of two points of the perfect set is the same as the relative order of the corresponding points in the continuum, the end-points of a complementary interval of the perfect set corresponding to one point of the continuum. To a closed set non-dense in H, there corresponds a closed set non-dense in the continuum; and a set of the first or second category relatively to H corresponds to a set of the first or the second category, respectively, in the continuum. It thus appears that the properties of sets of the first and of the second categories in the continuum, which have been above established, can be immediately extended to the case of sets of the first and second categories relatively to any perfect set H.

This is a particular case of the general property of any perfect set H considered as the domain in which sets of points are defined; viz. that H plays the same part relatively to such sets, as a continuous interval does relatively to sets of points defined in it.

EXAMPLES.

1. Let $P_1, P_2, P_3, \dots P_n, \dots$ be an enumerable set of points in an interval (a, b); the set may be everywhere-dense in (a, b). The finite sets

$$(P_1), (P_1, P_2), (P_1, P_2, P_3) \dots (P_1, P_2, \dots P_n) \dots$$

are each closed, and the given set is the limiting set, which is therefore of the first category. The remaining points of (a, b) form a set of the second category.

2. Denoting the points of the interval $(0, 1)$, as in Ex. 5, § 75, by

$$x = \frac{c_1}{k_1} + \frac{c_2}{k_1 k_2} + \dots + \frac{c_n}{k_1 k_2 \dots k_n} + \dots$$

where $c_n < k_n$; let the fixed integers $k_1, k_2, \dots k_n, \dots$ form a sequence which increases without limit. If* $a_1, a_2, \dots a_n, \dots$ is any sequence of positive integers which increase without limit, let G_n denote the set of those numbers x, which are such that the integers $c_1, c_2, \dots c_n, \dots$ are all $< a_n$. The sets $G_1, G_2, G_3, \dots G_n, \dots$ are a sequence of perfect sets, each one of which contains the preceding ones; the set G_ω is then a set of the first category.

3. The numbers of the continuum $(0, 1)$ may be divided into sets, of the first and the second categories, in the following manner:—All the numbers in $(0, 1)$ may be expressed as endless decimals; the finite decimals being therefore not used. Let† the set H consist of all those numbers in which the digit 9 occurs only a finite number of times, and of those numbers also in which, from and after some place, all the figures are 9. The complementary set K consists of all those numbers in which 9 occurs an infinite number of times, except those in which every figure is 9 from and after some place. The set H is the limit of a sequence of non-dense closed sets $H_1, H_2, \dots H_n, \dots$ each of which is of cardinal number c. For, let H_1 consist of the numbers of the form $\cdot abc \dots k999 \dots$, in which every figure is 9, after some fixed place, and in which none of the figures $a, b, c, \dots k$ is 9; together with those decimals in which no figure is 9. No number of the set H_1 can lie within the interval $(\cdot abc \dots k899 \dots, \cdot abc \dots k999 \dots)$ which is therefore a complementary interval of the set. The set H_n may be taken to consist of the numbers of the form $\cdot abc \dots hk999 \dots$, in which k is not 9, and not more than $n-1$ of the figures $a, b, c, \dots h$, are 9; together with those decimals in which 9 does not occur. That each of the sets H_n is of cardinal number c, follows from the fact that it contains all the decimals in which 9

* Brodén, *Math. Annalen*, vol. LI.

† See Schönflies, *Bericht über die Mengenlehre*, p. 106.

does not occur; and these, if interpreted in the scale of 9, represent all the numbers of the continuum (0, 1). The set H is everywhere-dense, since it contains that everywhere-dense set of numbers in which every figure is 9, after some place. The set K, being of the second category, is also everywhere-dense, and of cardinal number c.

4. The following method of dividing the continuum (0, 1) into two portions, each of which is everywhere-dense, and of cardinal number c, has been given by Brodén*:—Let $l_0+l_1+l_2+\ldots+l_n+\ldots$ denote a divergent series of positive numbers, such that the limit of l_n, as n is indefinitely increased, is zero. Let a be a positive number <1, and let $n_1, n_2, \ldots n_\iota, \ldots$ be a sequence of increasing positive integers. It is possible to choose the divergent series so that each of the ratios l_{n_2}/l_{n_1}, l_{n_3}/l_{n_2}, ... is $<a$: if this be done, the series $\sum_{\iota=1}^{\infty} l_{n_\iota}$ is convergent, its sum being $<\frac{l_{n_1}}{1-a}$. Each of the series obtained from $\sum_{\iota=1}^{\infty} l_{n_\iota}$, by leaving out a finite number of terms, is also convergent. The convergent series so obtained, form an unenumerable set: for they are obtained by multiplying the terms of the series $\sum_{\iota=1}^{\infty} l_{n_\iota}$, each either by 0, or by 1; and thus there is a series corresponding to each fractional number expressed in the dyad scale. Corresponding to each convergent series, there is a divergent series which consists of $l_0+l_1+\ldots+l_n+\ldots$, with the convergent series removed from it. We obtain in this manner an unenumerable set of divergent series. The convergent and divergent series, each of which consists of terms of $l_0+l_1+\ldots+l_n+\ldots$, may now be correlated with the numbers of the continuum (0, 1). Let these numbers be expressed in the dyad scale, in the form $\cdot a_1a_2a_3\ldots$, where every a is 0, or 1, and the case in which every figure is zero after some place, is excluded. To one of the series $l_p+l_q+l_r+\ldots$, we may take that number in which a_p, a_q, a_r, ... are all 1, and the remaining digits 0. The points of (0, 1) are thus divided into two classes; one of these consisting of all the numbers which correspond to convergent series, and the other of those corresponding to divergent series.

DIMINISHING SEQUENCES OF CLOSED SETS.

90. A closed set may be either

(1) a non-dense closed set defined, as we have shewn, as the end-points of an everywhere-dense set of non-overlapping intervals, together with the limiting points of the end-points, or

(2) a finite number of non-abutting closed intervals, or

(3) the set obtained by adding to a non-dense closed set the internal points of some of the complementary intervals.

The sets (2) may be regarded as the particular case of (3), which arises when the non-dense closed set is a finite one.

The closed sets here considered will be taken to be of any one of the types thus indicated.

Let $P_1, P_2, P_3, \ldots P_n, \ldots$ be an unending sequence of closed sets, each one of which contains the one which succeeds it; then it will be shewn that a

* *Crelle's Journal*, vol. CXVIII, p. 29.

set P_ω exists the points of which are contained in every one of the closed sets, and that *this set* $P_\omega \equiv D(P_1, P_2, \dots P_n, \dots)$ *is itself a closed set*, which may however contain only one point or a finite number of points.

To prove this theorem, suppose (a, b) divided into a finite number of parts; then in one at least of these parts (a_1, b_1) there must exist points which belong to P_n for every value of n; for otherwise the sequence would be a finite one. Dividing (a_1, b_1) into a finite number of parts, in one (a_2, b_2) at least of these there are points which belong to every P_n. Proceeding in this manner, and choosing the mode of division so that (a_n, b_n) converges to the limit zero, the point which is in the interior of all the intervals

$$(a_1, b_1), (a_2, b_2), \dots (a_n, b_n), \dots$$

is a point which belongs to every P_n, and is therefore a point of P_ω; thus P_ω contains one point at least.

To shew that P_ω is a closed set, let $p_1, p_2, \dots p_r, \dots$ be a sequence of points in it which has p for its limiting point. Then $p_1, p_2, \dots p_r, \dots$ are all points of P_n whatever n may be; and since P_n is closed, p is a point of P_n. This holds for every value of n, hence p is a point of P_ω; which establishes the result.

The theorem is a generalization of the results of § 63. In fact, if P_1, $P_2, \dots P_n, \dots$ are taken to be the derivatives $G^{(1)}, G^{(2)}, \dots G^{(n)}, \dots$ of a set G, the existence of the closed set $G^{(\omega)}$ follows from the theorem.

Again we may take $P_1 = G^{(\alpha_1)}, P_2 = G^{(\alpha_2)}, \dots P_n = G^{(\alpha_n)}, \dots$ where $\alpha_1, \alpha_2, \dots$ $\alpha_n, \dots$ is any sequence of numbers of the second class, of which β is the limiting number. The existence of the closed set $G^{(\beta)}$ then follows from the theorem.

91. Let us now suppose that, for the sequence $P_1, \dots P_n, \dots$ of closed sets, each one of which contains the next, a positive number C exists such that the content I_n of P_n is, for every n, greater than C. It will then be shewn* that *the content* I_ω *of* P_ω *is* $\geqq C$.

In order to establish this theorem, the following lemma is required:—

If G_1, G_2 be two closed sets with contents I_1, I_2 respectively, then the set $D(G_1, G_2)$ of points common to G_1, G_2 is a closed set of content I'; and the set $M(G_1, G_2)$ of points belonging either to G_1 or to G_2, or to both, is a closed set of content I'', where $I' + I'' = I_1 + I_2$.

That $D(G_1, G_2)$ is closed, follows from the fact that any limiting point of it must be a limiting point both of G_1 and of G_2, and therefore belongs to the set. That $M(G_1, G_2)$ is closed, follows from the fact that any limiting point of it must be a limiting point of one at least of the sets G_1, G_2. If the points

* This theorem was given by W. H. Young, in his paper on "Open sets and the theory of content," *Lond. Math. Soc. Proc.* Ser. 2, vol. II, p. 25.

of G_1 be removed from the set $M(G_1, G_2)$, the remainder is a measurable set of measure $I'' - I_1$, in accordance with the theorem of § 82. But the set so obtained could also be obtained by removing from G_2 those points which belong to $D(G_1, G_2)$; hence the measure of the set is also $I_2 - I'$. Therefore it follows that $I' + I'' = I_1 + I_2$.

Let us now suppose that, if possible, the set P_ω has its content I_ω less than C. Now the set $P_1 - P_\omega$ is measurable, and has for its measure $I_1 - I_\omega$; hence $P_1 - P_\omega$ contains as component a closed set Q_1 of content greater than $I_1 - I_\omega - \epsilon$, where ϵ is arbitrarily small, and is taken $< C - I_\omega$.

This closed set Q_1 has, in accordance with the lemma, in common with P_2, a closed component Q_2 whose content is $> I_2 + I_1 - I_\omega - \epsilon - J$, where J is the content of the set of all points belonging to the closed component Q_1 of P_1 only, or to P_2 only, or to both of these; and it is clear that $J \ngtr I_1$. Therefore the component Q_2 of P_2 has its content greater than the positive number $I_2 - I_\omega - \epsilon$, and is itself contained in the component Q_1 of P_1, of which the content is greater than $I_1 - I_\omega - \epsilon$. Proceeding in a similar manner, we obtain closed components $Q_3, Q_4, \ldots$ of $P_3, P_4, \ldots P_n, \ldots$, each of which contains the next, and none of which contains points of P_ω. Now these closed sets $Q_1, Q_2, \ldots Q_n, \ldots$ have, in accordance with the theorem of § 90, at least one point in common; hence $P_1, P_2, \ldots$ have a point in common which does not belong to P_ω; and this is contrary to the definition of P_ω.

THE COMMON POINTS OF A SYSTEM OF OPEN SETS.

92. *If* $G_1, G_2, \ldots G_n, \ldots$ be a sequence of sets of points, such that each set G_n contains the next G_{n+1}, and if the interior measure of each set is greater than some fixed number C, whatever n may be, then the set G_ω of points common to all the sets has an interior measure $\geqq C$.*

It will be observed that the sets are not assumed to be measurable.

Let $m_\iota(G_n)$ denote the interior measure of a set G_n; then, by § 81, closed components $P_1, Q_2, Q_3, \ldots Q_n, \ldots$ of the sets $G_1, G_2, \ldots$ can be found such that the content of P_1,

$$I(P_1) > m_\iota(G_1) - \frac{1}{2}\epsilon,$$

$$I(Q_2) > m_\iota(G_2) - \frac{1}{4}\epsilon,$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$I(Q_n) > m_\iota(G_n) - \frac{1}{2^n}\epsilon,$$

$$\cdots\cdots\cdots\cdots\cdots$$

* W. H. Young, *loc. cit.* p. 25.

where ϵ is an arbitrarily small positive number. The set Q_2 has a closed component P_2 which is also a component of P_1, of content

$$I(P_2) = I(Q_2) + I(P_1) - I\{M(P_1, Q_2)\} \geqq I(P_1) + I(Q_2) - m_i(G_1)$$
$$> I(Q_2) - \frac{1}{2}\epsilon$$
$$> m_i(G_2) - \left(\frac{1}{2} + \frac{1}{4}\right)\epsilon.$$

Next take that closed component of Q_3 which is also a component of P_2; it can be shewn as before that

$$I(P_3) > m_i(G_3) - \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8}\right)\epsilon,$$

and so on. We have now a sequence of closed sets $P_1, P_2, \dots P_n, \dots$ each of which contains the next, and such that the content of each of them is $> C - \epsilon$; therefore the set P_ω of points common to all these, has its content $\geqq C - \epsilon$, and P_ω is a component of G_ω. It follows, since ϵ is arbitrarily small, that the inner measure of G_ω cannot be less than C.

93. We are now in a position to establish the following general theorem:—

If $G_1, G_2, \dots G_n, \dots$ is a sequence of sets of points, each of which sets is a component of a closed set of finite content l, and if the interior measure of each of the sets $G_1, G_2, \dots G_n, \dots$ is greater than a fixed number C, then there exists a set of points of interior measure $\geqq C$, and of the power of the continuum, such that each point of the set belongs to an infinite number of the given sets.*

The conditions of the theorem are satisfied if all the sets lie in the same finite interval of length l; also the sets are not assumed to be measurable ones.

Choose a closed component of each of the given sets, of content $> C$; let these components be $Q_1, Q_2, \dots Q_n, \dots$. Choose an integer m such that $mC \leqq l < (m+1)C$, and let us consider the first $n (> m+1)$ of the sets $Q_1, Q_2, \dots$. The points common to any pair of these closed sets form a closed set, and the set which contains all the points which belong to at least two of the n closed sets is also a closed set $Q_{1,n}$ of content I_{1n}. Those points of $Q_{1,n}$ which belong to Q_1 form a closed set of content $\leqq I_{1n}$, hence there is a set of points of Q_1, of measure $\geqq I(Q_1) - I_{1n}$, which do not belong to any of the sets $Q_2, Q_3, \dots Q_n$; and the measure of this set is $> C - I_{1n}$. Similarly each of the sets $Q_2, Q_3, \dots Q_n$ has a component of measure $> C - I_{1n}$, consisting of points which do not belong to any of the other sets, or to Q_1. The measure of all

* W. H. Young, *loc. cit.* p. 25.

these sets added together is $> n(C - I_{1n})$; and it must be less than l, since the sets do not contain any points common to two of them, and they are all enclosed in a set of measure l. Hence

$$n(C - I_{1n}) < (m+1)C; \text{ or } I_{1n} > \left(1 - \frac{m+1}{n}\right)C.$$

It has thus been shewn that the closed set $Q_{1,n}$ has the power of the continuum, since its content is proved to be positive; and this holds for every value of n which is $> m+1$. Considering now the next n sets $Q_{n+1}, Q_{n+2}, \ldots Q_{2n}$, there is a closed set of content $> \left(1 - \frac{m+1}{n}\right)C$, consisting of points each of which belongs to two at least of the sets; and a similar result holds for each system of n sets $Q_{rn+1}, Q_{rn+2}, \ldots Q_{(r+1)n}$.

We have now an infinite sequence of closed sets $Q_{1,n}, Q_{2,n}, Q_{3,n}, \ldots Q_{r,n}, \ldots$ each of which has content $> \left(1 - \frac{m+1}{n}\right)C$, and the points of each of them belong to two at least of the given sets. By applying similar reasoning, and taking n' sets at a time, we see that there are an infinite number of sets each of content $> \left(1 - \frac{m+1}{n}\right)\left(1 - \frac{m+1}{n'}\right)C$, and such that each point of any one of them belongs to four at least of the given sets. Proceeding in this manner we obtain sets of points, each of content

$$> \left(1 - \frac{m+1}{n}\right)\left(1 - \frac{m+1}{n'}\right)\left(1 - \frac{m+1}{n''}\right)\ldots\left(1 - \frac{m+1}{n^{(s)}}\right)C,$$

and such that each point of each set belongs to at least 2^{s+1} of the given sets. Now let $n, n', \ldots n^{(s)}$ be so chosen, that

$$\frac{m+1}{n} < \frac{1}{2}\epsilon, \quad \frac{m+1}{n'} < \frac{1}{4}\epsilon, \ldots \frac{m+1}{n^{(s)}} < \frac{1}{2^{s+1}}\epsilon;$$

then the content of each of the sets which contains points belonging to 2^{s+1} at least of the given sets is

$$> C\left(1 - \frac{1}{2}\epsilon\right)\left(1 - \frac{1}{4}\epsilon\right)\ldots\left(1 - \frac{1}{2^{s+1}}\epsilon\right) > C\left\{1 - \epsilon\left(\frac{1}{2} + \frac{1}{4} + \ldots + \frac{1}{2^{s+1}}\right)\right\}$$
$$> C(1-\epsilon).$$

The process can be carried on without limit; and we see that the set which consists of all points belonging to 2^{s+1} at least of the given sets contains closed components of content $> C(1-\epsilon)$. Considering the sequence $P_1, P_2, \ldots$ of sets such that P_1 contains all points that belong to two at least of the given sets, P_2 contains all points that belong to 2^2 at least of the given sets, and so on, it is clear that P_1 contains P_2, and P_2 contains P_3, etc. But the interior measure of each set is $> C(1-\epsilon)$; hence, in accordance with the theorem of § 92, there exists a set of points common to all the sets

$P_1, P_2, \ldots$ of interior measure $\geqq C(1-\epsilon)$. This set consists of points which belong each to an infinite number of the given sets; and its interior measure is $\geqq C$, since ϵ is arbitrarily small. The set has the power of the continuum, since it contains closed components of content greater than zero.

The theorem that has been now established is of considerable importance on account of the applications of it which can be made in various parts of the theory of functions; it is due* to W. H. Young. That particular case of the theorem in which the sets are all measurable was first stated†, without proof, by Borel.

An important case of the theorem arises if we suppose each of the sets to consist of a finite, or an enumerably infinite, set of closed intervals; in which case the sets are all measurable. The theorem may then be stated as follows:—

If there be given an infinite number of sets of intervals, in a finite segment, each set consisting of a finite, or enumerably infinite, number of non-overlapping intervals, and if the measure of each set of intervals is greater than some fixed positive number C, then there exists a set of points having the power of the continuum, and of interior measure $\geqq C$, such that each point of the set belongs to an infinite number of the given sets of intervals.

This theorem contains the completion, and generalization, of a theorem due to Arzelà‡ which is stated by him as follows:—

Let y_0 be a limiting point of any set of numbers (y), and let

$$G_0 = (y_1, y_2, \ldots)$$

be a sequence of numbers of (y) which converges to the limit y_0. Assuming the variables to be orthogonal coordinates of a point in a plane, let the set of straight lines $y = y_1, y = y_2, y = y_3, \ldots$, be drawn, and let a set of intervals be taken on the portion of each of these straight lines which is in the interval (a, b) of x. Suppose that each set of intervals is finite in number, and that this number is variable from one straight line to another, but increases indefinitely as the index in y_s increases indefinitely. Let the sum of the intervals $\delta_{1,s}, \delta_{2,s}, \ldots \delta_{n,s}$ on the line $y = y_s$, be d_s. If for every value of s, d_s is greater than C, a determinate positive number, there necessarily exists at least one point x_0 in the interval (a, b), such that the straight line $x = x_0$ intersects an infinite number of the intervals δ.

Arzelà subsequently removed the condition that each set of intervals is a finite one.

* *Proc. Lond. Math. Soc.* Ser. 2, vol. II, p. 26.

† *Comptes Rendus*, December 1903.

‡ *Rend. dell' Ac. dei Lincei* (4) 1, (1885), p. 637; a second proof, which is however not rigorous, has been given by Arzelà in the *Memorie della R. Acc. d. Sc. di Bologna*, Ser. 5, vol. VIII, 1899.

THE ANALYSIS OF SETS IN GENERAL.

94. It has been proved that a closed set can always be analysed into an enumerable set and a perfect one, either of which may be absent; we now proceed to consider the case of a set which is not necessarily closed. Before doing so, it is necessary to classify the points of a set, according to the cardinal number of those points of the set which are contained in the immediate neighbourhoods of such points*.

An isolated point of a set G is such that in a sufficiently small neighbourhood of the point there are no other points of G. For this reason *an isolated point may be said to be of degree zero in the set.*

A point P, which is a limiting point of G, and is such that in a sufficiently small neighbourhood of P there is an enumerable set of points of G, is said to be a *point of enumerable degree in the set*, or of *degree a in the set.*

If a limiting point P of the set G is such that in any neighbourhood of P, however small, there is an unenumerable set of points of G, the point P is said to be of *unenumerable degree in the set G.*

In case the point P is such that, in every neighbourhood of it, the cardinal number of the points of G contained in such neighbourhoods is c, the point is said to be of *degree c in the set.*

It is not definitely known whether cardinal numbers exist which lie between a and c; but if any such cardinal number x exists, a point would be of degree x in the set, if a neighbourhood of P exists such that in that neighbourhood, and in every smaller neighbourhood, there is contained a part of the set which is of cardinal number x.

The points of unenumerable degree consist of all the points whose degrees in the set are greater than a.

If a set G contains no point which is of unenumerable degree in the set, the set is enumerable or finite.

If P be any point of the set, an interval which contains P can be found, such that the part of G contained in this interval is enumerable; and the same holds for any point Q of G which is not contained in the interval round P. In this manner we can proceed until we have a non-overlapping set of intervals which contain all the points of G. Since these intervals form an enumerable or a finite set, and in each of them there is a finite or enumerable part of G, it follows that G is an enumerable set.

* The analysis here carried out, of sets in general, was given by Cantor, *Acta Math.* vol. VII. A more elementary presentation of the subject has been given by W. H. Young, *Quart. Journ. of Math.* vol. XXXV, 1903.

$P_1, P_2, \ldots$ of interior measure $\geqq C(1-\epsilon)$. This set belong each to an infinite number of the given sets: is $\geqq C$, since ϵ is arbitrarily small. The set has th since it contains closed components of content gre

The theorem that has been now established is on account of the applications of it which can the theory of functions; it is due* to W. H. of the theorem in which the sets are all measur proof, by Borel.

An important case of the theorem arises to consist of a finite, or an enumerably in which case the sets are all measurable. T follows:—

If there be given an infinite numb segment, each set consisting of a finite, o overlapping intervals, and if the meas than some fixed positive number C, the power of the continuum, and of inter the set belongs to an infinite number o

This theorem contains the comp to Arzelà‡ which is stated by him

Let y_0 be a limiting point of

be a sequence of numbers of (the variables to be orthogonal straight lines $y = y_1, y = y_2, \ldots$ be taken on the portion of (a, b) of x. Suppose that this number is variable indefinitely as the index intervals $\delta_{1,s}, \delta_{2,s}, \ldots \delta_{n,s}$ d_s is greater than C, a d at least one point x_0 in intersects an infinite nu

Arzelà subsequent finite one.

* *Proc. Lond. Math.*
† *Comptes Rendus,*
‡ *Rend. dell' Ac.*
has been given by Ar

dense- rhood of P neighbourhood se cannot all be numerable set of such ghbourhood of P, there degree in G; therefore P which are of unenumerable

set G which is unenumerable An enumerable set may or -in-itself; but if it does contain a perfect set, and thus the enu-

established that, *if G is an unenumer- of the same degree $x\,(\geqq a)$ in G cannot exceed x, and which is dense-*

P_x in an interval such that x is the interval. Doing the same with any is not in the interval round P, and in intervals, we have a finite, or intervals, in each of which is a set of all the points of degree x are included the general theory of cardinal numbers, points of G included in this enumerable hence the set of points P_x cannot have a

-in-itself, we observe that, in a sufficiently there is a set of points of G, of cardinal can be of degree in G higher than x. of degree higher than x, in some interval

Q in its interior, there would be
than x; but this is impossible,
, the points of G in (α, β) cannot
y were so, their cardinal number
rary to the hypothesis. Moreover,
ourhood of P_x there are points of the
ng point for such points. Therefore the
-in-itself.

e is dense-in-itself, and of cardinal number c.

ts of isolated points which form an enumerable
of G, and of limiting points which form a set called
Denoting the adherence and the coherence of G by Ga,
have $G = Ga + Gc$.

an in a similar manner be split up into its adherence and its
ich we denote by Gca and Gc^2 respectively; thus

$$Gc = Gca + Gc^2.$$

Gca is an isolated set, and therefore enumerable; and if we proceed
solve Gc^2 in a similar manner into its adherence Gc^2a, and its coherence
, and then to resolve Gc^3, it is clear that the process may be continued
any number n of times. We thus obtain

$$G = Ga + Gca + Gc^2a + \ldots + Gc^{n-1}a + Gc^n.$$

The set $Gc^{n-1}a$ may be named the adherence of G of order n, and Gc^n may be denominated the coherence of G of order n.

It may happen that, for some value of n, Gc^n vanishes; in that case G has been split up into a finite number of enumerable sets, and is consequently itself enumerable. If this be not the case, the process may be continued indefinitely, and Gc^n then exists for every value of n. We then define

$$D(G,\ Gc,\ Gc^2,\ \ldots\ Gc^n,\ \ldots),$$

the set of points common to all the coherences of G, to be the coherence of order ω, and denote it by Gc^ω. It is clear that every point of G which does not belong to one of the sets $Gc^{n-1}a$, belongs to Gc^ω, hence we have

$$G = \Sigma Gc^{n-1}a + Gc^\omega,$$

the summation being taken for all values of n.

We now split up Gc^ω into its adherence $Gc^\omega a$, and its coherence $Gc^{\omega+1}$, and proceed further to obtain the adherences and coherences of G of the orders of the various numbers of the second class. If $\alpha_1, \alpha_2, \ldots \alpha_n, \ldots$ is a sequence of numbers of the second class, which has β for its limit, the coherence of order β is defined by

$$Gc^\beta = D(Gc^{\alpha_1},\ Gc^{\alpha_2},\ \ldots\ Gc^{\alpha_n},\ \ldots).$$

We now obtain a resolution of G of the form

$$G = \Sigma G c^p a + G c^\gamma,$$

where γ is any number of the first or second class, and the summation refers to all values of p which are less than γ. Each adherence $Gc^p a$ is an isolated set, and therefore enumerable; and if G contains a component which is dense-in-itself, this component is contained in Gc^γ.

First suppose G to be an enumerable set; the process of analysis must then cease for some number γ of the first or second class. For if $Gc^p a$ existed for every number γ of the second class, we should have obtained an unenumerable set of adherences containing no points in common, and all belonging to G: thus G could not be enumerable.

The cessation of the process may take place in two different manners:—

(1) if for some number γ of the first or second class, $Gc^\gamma \equiv 0$, G has been resolved into an enumerable set of adherences, and it contains no component which is dense-in-itself:

(2) if for some number γ,

$$Gc^\gamma a = 0,$$

in which case $Gc^\gamma = Gc^{\gamma+1}$, the set Gc^γ then contains no adherence, and every point of it is a limiting point, and Gc^γ is therefore dense-in-itself. The set G has consequently been resolved into an enumerable component which contains no part that is dense-in-itself, and into a set which is enumerable and dense-in-itself.

Next, let us suppose that G is an unenumerable set. Then it has been shewn that those points of G which are of unenumerable degree in G form a set that is dense-in-itself; and those points which belong to the adherences of all orders are points of zero, or of enumerable, degree, and thus form an enumerable set. It follows, since all points that do not belong to that part of G which is dense-in-itself belong to the adherences, that the number of adherences must be enumerable; and thus that, for some number γ of the first or second class, Gc^γ is dense-in-itself. The set Gc^γ may consist of an enumerable set dense-in-itself, and of sets of higher cardinal numbers dense-in-themselves.

It has thus been shewn that *any set G may be represented by*

$$G = U + V_a + \Sigma V_x + V_c,$$

where U is an enumerable set which contains no component that is dense-in-itself, V_a is an enumerable set of points of degree a dense-in-itself, V_c is a set of cardinal number c consisting of points of degree c dense-in-itself, V_x is a set dense-in-itself consisting of points of degree x, where $a < x < c$.

If, as is probable, no cardinal numbers exist between a and c, the sets V_x can be omitted. A set such as V_a, V_x, V_c is denominated a homogeneous set of degree a, x, c, in the set G.

If G is a closed set, then as has been shewn in § 73, V_c is perfect, and ΣV_x cannot exist.

INNER LIMITING SETS.

96. Let us suppose that each rational point $\frac{p}{q}$ in the interval (0, 1) is enclosed in the interval $\left(\frac{p}{q}-\frac{\lambda}{q^3}, \frac{p}{q}+\frac{\lambda}{q^3}\right)$, where λ has the same value for all the points. In this manner the rational points are enclosed in a set of overlapping intervals, whose sum is less than $\lambda\Sigma\,(q-1)\frac{2}{q^3}$, or than $2\lambda\Sigma\frac{1}{q^2}$, which can be made as small as we please by choosing λ small enough. The equivalent set of non-overlapping intervals defines, by means of the end-points and their limits, a closed set $\{q_1\}$, such that for any point of the set $\left|\frac{p}{q}-q_1\right| \geq \frac{\lambda}{q^3}$, for all points $\frac{p}{q}$.

Now consider the set of points defined by

$$x=\frac{a_1}{10}+\frac{a_2}{10^{2!}}+\frac{a_3}{10^{3!}}+\ldots+\frac{a_n}{10^{n!}}+\ldots,$$

where each a is ≤ 9, and the a are such that an infinite number of them are different from zero. It has been shewn by Liouville that these numbers x are transcendental. Let

$$\frac{p}{q}=\frac{a_1}{10}+\frac{a_2}{10^{2!}}+\ldots+\frac{a_n}{10^{n!}}, \text{ thus } q=10^{n!},$$

then
$$x-\frac{p}{q}=\frac{a_{n+1}}{10^{(n+1)!}}+\ldots<\frac{1}{q^n}\left(\frac{a_{n+1}}{q}+\ldots\right)<\frac{1}{q^n}.$$

It follows that, if x is one of the above transcendental numbers, whatever value λ may have, it is interior to an interval $\left(\frac{p}{q}-\frac{\lambda}{q^3}, \frac{p}{q}+\frac{\lambda}{q^3}\right)$. For suppose $q=10^{n!}$; then $\left|\frac{p}{q}-x\right|<\frac{1}{q^n}<\frac{\lambda}{q^3}$, provided $\lambda \geq \frac{1}{10^{(n-3)\,n!}}$; and, however small λ may be, values of n can be found for which this inequality is satisfied. Therefore rational points $\frac{p}{q}$ can be found however small λ may be, such that x lies within the intervals $\left(\frac{p}{q}-\frac{\lambda}{q^3}, \frac{p}{q}+\frac{\lambda}{q^3}\right)$. It thus appears that, besides the original points $\frac{p}{q}$ enclosing which the intervals are drawn, there are other points which lie inside the intervals for all values of λ, when λ is diminished indefinitely.

This example, which is due to Borel*, shews that if each point x of a set be enclosed in a series of intervals $\delta_1(x), \delta_2(x), \ldots \delta_n(x), \ldots$ assigned according to some prescribed law, and such that the upper limiting value of $\delta_n(x)$ for all the points of $\{x\}$ has the limit zero when n is increased indefinitely, the magnitude and position of $\delta_n(x)$ being assigned for each n and each x, then, in general, there are points which do not belong to the given set $\{x\}$ remaining in the interior of the set of intervals $\{\delta_n(x)\}$, however great n may be. It is clear that every point x' not belonging to $\{x\}$, which is in the interior of an interval of the set $\{\delta_n(x)\}$, for every value of n, must be a limiting point of the set $\{x\}$. For if p is a point which is not a limiting point of $\{x\}$, the distances of p from the points of the set have a definite finite minimum c; hence, when n is so great that the upper limit of $\delta_n(x)$, for a fixed value of n, and for all points of $\{x\}$, is less than c, the point p is exterior to all the intervals of the set $\{\delta_n(x)\}$.

The points not belonging to the unclosed set $\{x\}$ which are interior to the set $\{\delta_n(x)\}$ for every value of n, are among the limiting points of the set $\{x\}$: and it will appear that some or all of the limiting points of $\{x\}$ may have this property, according to the law of choice of the intervals of the set.

If all the intervals $\{\delta_n(x)\}$ be taken of equal length $2c_n$, with the x in the centre of its interval, where c_n has the limit zero when n is indefinitely increased, then every limiting point of $\{x_n\}$ lies within the set $\{\delta_n(x)\}$. For, however small c_n may be, there are points of the set whose distance from a limiting point p is less than c_n.

If the points of a set G be enclosed in a series of sets of intervals $\{\delta_n(x)\}$, which are subject to the condition that the maximum of the lengths $\delta_n(x)$ for all points of the set has the limit zero, when n is increased indefinitely, then the set G, together with those points, if any, of the derivative G', not being points of G, which are within the intervals $\{\delta_n(x)\}$ for every value of n, is said to be the inner limiting set† *for the sequence $\{\delta_n(x)\}$ of sets of intervals.*

If G is a given set of points, and it is possible so to choose the sequence of sets $\{\delta_n(x)\}$, that no points which do not belong to G remain in the interior of the intervals $\{\delta_n(x)\}$ for every value of n, then the set G is said to be an inner limiting set of points.

It has been shewn above that every closed set is an inner limiting set.

* *Leçons sur la théorie des fonctions*, p. 44.

† This term is due to W. H. Young, who has investigated the properties of such sets, see *Leipziger Berichte*, August 1903, "Zur Lehre der nicht abgeschlossenen Mengen." For further properties see also *Proc. Lond. Math. Soc.* Ser. 2, vol. I, p. 262.

If a point which does not belong to an inner limiting set is interior to one or more of the intervals $\{\delta_{n-1}(x)\}$, but is not interior to any of the intervals $\{\delta_n(x)\}$, then that point will be said to be *shed* from the sequence of sets of intervals at the index n.

In accordance with the theorem of § 66, the set $\{\delta_n(x)\}$ may be replaced by a set of non-overlapping intervals $\{\Delta_n\}$ which have the same internal points as $\{\delta_n(x)\}$; and the points which are not interior to $\{\Delta_n\}$ or to $\{\delta_n(x)\}$ form a closed set. It thus appears that *every inner limiting set is complementary to a set of points which is the limit of a sequence G_n of closed sets, such that G_n is contained in G_{n+1}*. Conversely, *every set which is complementary to G_∞, the limit of an ascending sequence of closed sets, is an inner limiting set*, the intervals complementary to G_n being taken as $\{\Delta_n\}$.

In case the closed sets G_n are all non-dense, the set G_∞ is a set of the first category, and the complementary set is of the second category. Therefore it follows that *every set of the second category is an inner limiting set.*

Every enumerable set is a set of the first category, for it may be exhibited as the limit of a sequence of finite sets; hence the complementary set is of the second category, and is therefore an inner limiting set.

In the case of any enumerable set $\{P\}$, those of its limiting points which do not belong to $\{P\}$ form an inner limiting set.

To prove this, let Q be the set of those limiting points of the set $P_1, P_2, \ldots P_n, \ldots$ which do not belong to that set; then the points of Q can be enclosed in intervals $\{\delta_1\}$ which do not contain P_1; and in the interior of the intervals $\{\delta_1\}$ a set $\{\delta_2\}$ may be chosen enclosing the set Q, and excluding the point P_2, and so on. Then the sequence of sets $\{\delta_n\}$ has for its inner limiting set $\{Q\}$; and the only limiting points of $\{Q\}$ which do not belong to $\{Q\}$ are, or may be, the points $P_1, P_2, \ldots P_n, \ldots$ which have each been shed at a definite index.

It can easily be seen that an inner limiting set remains such, if a finite number of points be added to, or subtracted from the set. Also the sum of a finite number of inner limiting sets is itself an inner limiting set; but this is not in general true of the sum of an indefinitely great number of inner limiting sets.

97. It will now be shewn that *every inner limiting set is either enumerable or else of the power of the continuum.*

Let an inner limiting set P be defined by means of a sequence of sets $\{\Delta_n^{(1)}, \Delta_n^{(2)}, \ldots \Delta_n^{(r)}, \ldots\}$, each of which consists of non-overlapping intervals. The set $\{\Delta_n\}$ is measurable, and its measure is

$$\underset{r=\infty}{\mathrm{L}} \{\Delta_n^{(1)} + \Delta_n^{(2)} + \ldots + \Delta_n^{(r)} + \ldots\} = m_n,$$

where m_n diminishes as n increases. If m_n has a limit, when n is indefinitely increased, which is greater than zero, say C, then $m_n > C$, for every value of n; and thus, in accordance with the theorem of § 92, the inner limiting set has a measure $\geqq C$, and therefore contains closed components of positive content; therefore in this case the inner limiting set has the power c of the continuum.

There remains now to consider the case in which m_n has the limit zero, when n is increased indefinitely. It is clear that in this case no interval of $\{\Delta_n\}$ can also be an interval of all the sets $\{\Delta_{n+1}\}$, $\{\Delta_{n+2}\}, \ldots$; for, if it were so, the measure of all these sets would exceed the length of the particular interval, which is contrary to the hypothesis $\mathrm{L}\, m_n = 0$. Let us first suppose that the sets $\{\Delta_n\}$ are all everywhere-dense in the interval in which they are all contained; then any particular interval $\Delta_n^{(r)}$, since it cannot be an interval of all the following sets, must contain at least two intervals of one of the following sets $\{\Delta_{n'}\}$. Let us denote these two intervals by d_0, d_1. Applying the same argument to each of the intervals d_0, d_1, each must contain two intervals of some following set; and thus we have four intervals d_{00}, d_{01}, d_{10}, d_{11}, all contained in $\Delta_n^{(r)}$. Proceeding in this manner, we have intervals d with indices consisting of every permutation of the digits 1 and 0; and if we consider any sequence, such as d_{01}, d_{011}, d_{0110}, $d_{01100}, \ldots$, the indices form a sequence of radix fractions expressed in the dyad scale, each interval containing the next in its interior; for it is clear that at each stage of the process the intervals in a particular interval may be so chosen that they have no end-point in common with it. Since a sequence of intervals can thus be found which corresponds to any irrational fraction expressed in the dyad scale, and since there must be a point of P in the interior of all the intervals of such sequence, it appears that in $\Delta_n^{(r)}$ there is a set of points of P which has the same power as the set of irrational numbers between 0 and 1; and that power is c.

Next, let us suppose that the sets of intervals $\{\Delta_n\}$ are not all of them everywhere-dense in their domain; and suppose that the inner limiting set P contains a part Q which is dense-in-itself, so that the derivative Q' is perfect. The perfect set Q' may be placed into correspondence with all the points of the continuum (0, 1) so that the order of corresponding points in the two sets is the same; and to each point in the second continuum there corresponds a single point of Q', except that the end-points of an interval complementary to Q' correspond to a single point in the second continuum. The points of Q correspond to points of a set Q_1 everywhere-dense in the second continuum; and those intervals of the set $\{\Delta_n\}$ which contain points of Q, correspond to intervals of a set $\{\Delta_n'\}$, which is everywhere-dense in the second continuum.

Those points of the second continuum which are interior to all the sets

$\{\Delta_n'\}$, form a set of power c, as has been shewn above; it therefore follows that the set Q has also the power of the continuum. For the points interior to all the sets $\{\Delta_n'\}$ are all either points of Q_1, or else points which correspond to the end-points of intervals complementary to Q', and these latter form at most an enumerable set.

The following theorem has now been established:—

An inner limiting set has the power of the continuum if it contains a component which is dense-in-itself, and if it contains no such component it must be enumerable. Its measure is the lower limit of the measures of the non-overlapping set of intervals by which it is defined.

For the only sets which contain no component dense-in-itself are enumerable; and it has been shewn that an inner limiting set which contains such component has the power of the continuum.

It thus appears that *an enumerable set, which contains a component which is dense-in-itself, cannot be an inner limiting set.*

98. It will now be shewn that *every enumerable set, which contains no component that is dense-in-itself, is an inner limiting set.*

Let P be an enumerable set, and let us first suppose that the derivative P' is also enumerable; then in this case P contains no component dense-in-itself, for the derivative of such a component would be perfect, and would be a component of P', which is impossible when P' is enumerable. Divide P into two parts P_1 and P_2; and of these let P_1 consist of those points which are not limiting points of the set $P' - D(P, P')$, composed of those points of P' which do not belong to P; while the other part P_2 consists of those points which are limiting points of $P' - D(P, P')$. Since all the limiting points of the enumerable set $P' - D(P, P')$ which do not belong to the set itself belong to P_2, the set P_2 is, as has been shewn in § 96, an inner limiting set. The points of P_1 not being limiting points of $P' - D(P, P')$, each point of P_1 can be enclosed in an interval which contains no points of $P' - D(P, P')$; and the set of intervals thus obtained can be taken as the set $\{\delta_1\}$ of intervals enclosing P_1. It follows that, since the points of $P' - D(P, P')$ are not contained in a properly chosen sequence of sets of intervals enclosing the points of P_1, and are each shed at a definite index from a properly chosen sequence of intervals enclosing the points of P_2, the set $P_1 + P_2$ or P is an inner limiting set. We have now shewn* that:—

Every reducible set is an inner limiting set.

Next let us suppose that P', the derivative of the enumerable set P, has the power of the continuum. If P' contained all the points of any interval (α, β), P could not be an inner limiting set; for the points of P in (α, β) would

* See Hobson, *Proc. Lond. Math. Soc.* Ser. 2, vol. II.

be everywhere-dense in this interval, and would form a set dense-in-itself, which has been shewn to be impossible. Since P' does not contain all the points in any interval, and is closed, it can be resolved into the sum of a perfect set G_1 and an enumerable set L_1, consisting of points interior to the intervals complementary to G_1. The set P may be divided into two parts P_1 and Q_1, where P_1 consists of those points which are interior to the complementary intervals of G_1, and Q_1 consists of those points which belong to G_1: it may happen that Q_1 does not exist. It can be shewn that P_1 is an inner limiting set, whether Q_1 exists or not. For P_1 consists of a series of sets $P_{11}, P_{12}, \dots P_{1n}, \dots$ interior to the complementary intervals $(a_1, b_1), (a_2, b_2), \dots (a_n, b_n), \dots$ of G_1; but the set P_{1n} in (a_n, b_n) has all its limiting points in that interval, and those belonging to L_1 are enumerable; and therefore, in view of what has been proved above, P_{1n} is an inner limiting set.

The sequence of sets of intervals which enclose the points of P_{1n} may be so chosen that all the intervals of every set are interior to (a_n, b_n); thus no limiting points of P not belonging to P_1, except those belonging to P_{1n}', are ever interior to any interval of the sequence assigned to P_{1n}; and as this holds for every n, it follows that P_1 is an inner limiting set, and its points are such that they can be enclosed in a sequence of sets of intervals which from the beginning contain no points of G_1.

The set Q_1 consists of points which belong to G_1, and therefore Q_1 has no limiting points in L_1. If every point of G_1 were a limiting point of Q_1, the set Q_1 being dense in G_1, would be dense-in-itself; were it so, Q_1 could not be an inner limiting set. It follows that Q_1 is not dense in G_1, and thus Q_1' does not contain all the points of G_1. Let Q_1' be resolved into an enumerable set L_2 and a perfect set G_2: the latter may be absent. The set Q_1 may then be resolved into a component P_2 contained in the intervals complementary to G_2, and a component Q_2 contained in G_2; thus $P = P_1 + P_2 + Q_2$. The same argument applied to P_2, as was applied to P_1, shews that P_2 is an inner limiting set; and the intervals of the sequence which encloses its points may be taken to be all interior to the complementary intervals of G_2. The set Q_2 in G_2 may be treated as Q_1 in G_1 was treated, and we thus have $Q_2 = P_3 + Q_3$, where P_3 is an inner limiting set, and Q_3 is contained in a perfect set G_3. Proceeding in this manner, it may happen that for some integer n, Q_n does not exist, and then P is expressed as the sum of a finite number n of inner limiting sets, and is itself therefore an inner limiting set. If no integer n exists for which this happens, we consider the set $M(P_1, P_2, \dots P_n, \dots)$, where n has every integral value. It may happen that this set contains every point of P; but if not, we take the set

$$P - M(P_1, P_2, \dots P_n, \dots),$$

and resolve it as before into an inner limiting set P_ω, and a set Q_ω contained in a perfect set G_ω, but which cannot be dense in G_ω, since it cannot be dense-in-itself. We then proceed to resolve Q_ω into $P_{\omega+1}$ and a set $Q_{\omega+1}$ contained in a perfect set $G_{\omega+1}$. We proceed further, and may obtain in this manner sets whose index is any transfinite ordinal number of the second class; and thus P is resolved into

$$P_1 + P_2 + \ldots + P_\omega + P_{\omega+1} + \ldots + P_\beta + Q_\beta,$$

where β is a non-limiting number of the second class, or else into

$$P_1 + P_2 + \ldots + P_\omega + \ldots + P_\beta + \ldots$$

with no last term. Since P is enumerable, this process must come to an end at, or before, some definite number α of the second class; and the end can only come, either when there is no component Q_α in G_α, or when there is no G_α.

It has thus been shewn that, when P contains no component that is dense-in-itself, it can be resolved into a finite, or enumerably infinite, set of inner limiting sets, of which there may, or may not, be a last set. Let P_γ be one of the components into which P has been resolved, γ denoting a number of the first or second class. We now fix on a sequence of sets of intervals enclosing the points of P_γ, such that all the intervals are interior to the intervals complementary to G_γ; then the set $P_{\gamma+1} + P_{\gamma+2} + \ldots$, which is contained in G_γ, has no limiting points in any of the intervals which enclose the points of P_γ, for all its limiting points must be in G_γ. The sequence of sets of intervals having thus been fixed for every P_γ, we can now shew that each limiting point p of P, which does not belong to P, is shed from the whole sequence of sets of intervals, at a definite index. The point p is either a limiting point of P_1, belonging to L_1, or is contained in G_1. In the former case it is shed from the intervals enclosing P_1 at a definite index; and, not being a limiting point of $P_2 + P_3 + \ldots$, it is shed from the intervals enclosing the points of that set, at a definite index; consequently it is shed from the intervals enclosing P, at a definite index, the greater of the two former ones. In the latter case, unless p is in G_2 or in P_2', it is not a limiting point of $P_2 + P_3 + \ldots$, and never comes into any of the intervals enclosing the points of P_1; it is therefore shed at a definite index. If p belongs to $G_1, G_2, \ldots$ and to every G before G_α, but is not in G_α, it may be a point of P_α'. In that case it is not a limiting point of the set $P_{\alpha+1} + P_{\alpha+2} + \ldots$, and does not come into the interior of any of the intervals which enclose the points of $P_1, P_2, \ldots$, or any P with index less than α. It is therefore shed, at a definite index, from the sequence of sets of intervals enclosing the points of P. It has thus been established that:—

The necessary and sufficient condition that an enumerable set may be an inner limiting set is that it contains no component which is dense-in-itself.

A corollary to the above proof is that every enumerable set is the sum of an inner limiting set, and of a set which is dense-in-itself.

99. Any unenumerable set can, in accordance with the result of § 95, be expressed in the form $P = U + V_a + \Sigma V_x + V_c$; and we observe that if V_c is absent, the necessary and sufficient conditions that P may be an inner limiting set are that V_a and ΣV_x should both be absent; this follows from the preceding results.

If V_c exists, we observe that no point of $U + V_a + \Sigma V_x$ can be a limiting point of V_c; for any limiting point of V_c must be a point of degree c in the set P. If V_c is everywhere dense in (a, b) it follows that $U + V_a + \Sigma V_x$ is absent. The set V_c may be non-dense in (a, b), or it may be dense in some parts of (a, b) and non-dense in other parts.

It will be shewn that V_c is in general made up of a part which is non-dense in (a, b) and of a finite, or indefinitely great, number of parts each of which is everywhere-dense in a particular interval in which it lies. Suppose that an interval (α, β) can be found in which V_c is everywhere-dense; and let x be a point in (a, b) such that $x \geqq \beta$. Then those values of x for which V_c is everywhere-dense in (α, x), together with those values for which this is not the case, define a section of all the numbers of the continuum (β, b); and this section defines a number $\beta_1 \geqq \beta$. Similarly we may assign a number $\alpha_1 \leqq \alpha$, so that (α_1, β_1) is the greatest interval containing (α, β) which is such that V_c is everywhere-dense in it. If, in the parts of (a, b) external to (α_1, β_1), the set V_c is dense in any interval, then we proceed to fix the greatest interval for which it is everywhere-dense. In this manner we obtain a finite, or enumerably infinite, set of detached intervals contained in (a, b), in each of which V_c is everywhere-dense; and the remainder of (a, b) may consist of a set of detached intervals and of a set of points. In this remainder the points of V_c form a non-dense set.

No point of $U + V_a + \Sigma V_x$ can be in an interval (α_1, β_1) in which V_c is everywhere-dense. If $\overline{V}_c$ is the part of V_c which is non-dense in (a, b), every point of $U + V_a + \Sigma V_x$ must lie in one of the intervals complementary to the perfect set $\overline{V}_c'$. It is to be observed that in $\overline{V}_c$ are included the end-points of the intervals (α_1, β_1), in case those end-points belong to V_c.

In order that P may be an inner limiting set, it is necessary that the part of $U + V_a + \Sigma V_x$, which is in each interval complementary to $\overline{V}_c'$, should be an inner limiting set; and this cannot be the case unless V_a and ΣV_x are absent.

It has thus been shewn that:—

In order that an unenumerable set of points may be an inner limiting set, it is necessary that the set should contain no points whose degrees in the set are other than 0, *a, or c, and that it should contain no component which is dense-in-itself, and of which the points are of degree a in the set.*

The determination of the necessary and sufficient conditions that any given unenumerable set of points, however defined, may be an inner limiting set has now been reduced to the problem of determining the criteria for the case of a set which is dense-in-itself and all the points of which are of degree c in the set. The case in which the latter set is non-dense in its domain may be reduced, by the method of correspondence, to that in which it is everywhere dense; and the problem is therefore reducible to that of determining the conditions under which a given everywhere-dense set of points all of degree c in the set may be a set of the second category. No investigation of all the possible types of such sets has yet been carried out, and therefore the problem remains as yet unsolved. A set which is everywhere-dense in (a, b), and of which the points in every sub-interval have the power of the continuum, may be of the first category, and thus not be an inner limiting set; or it may be of the second category, and therefore be an inner limiting set. The question has been raised by Schönflies* whether every such set is necessarily either of the first or of the second category; this question must certainly be answered in the negative. For, if we divide (a, b) into any finite number of parts, and place in them alternately inner limiting sets which are dense and of power c, and dense sets of the first category and of power c, it is clear that the whole set so constituted cannot be either of the first or of the second category. The outstanding question as to the criteria that such sets may be of the second category, is of considerable importance in relation to the Theory of Functions.

NON-LINEAR SETS OF POINTS.

100. Most of the properties of linear sets of points can be extended without essential modification to the case of sets of points in two, three, or more dimensions; and those respects in which sets of points in more than one dimension differ, as regards the formulation of their properties, from linear sets are sufficiently exemplified by the case of plane sets. It will therefore be sufficient, for the purpose of indicating the principal properties of non-linear sets, to confine our account to the case of plane sets.

Each point (x, y) of a plane set is defined by the two numbers x, y which are the rectangular Cartesian coordinates of a point. A set which extends over the whole plane may be made to correspond with the points of a set which lies in a finite rectangle; this correspondence may be made by means of the relations $x = \tan\frac{\pi X}{2}$, $y = \tan\frac{\pi Y}{2}$, when X, Y are each restricted to have values between $+1$ and -1. We shall consequently assume that the plane sets under consideration consist of points lying in a finite rectangle whose sides are parallel to the axes of coordinates.

* See Schönflies, *Göttinger Nachrichten*, 1899, p. 282, also *Bericht über die Mengenlehre*, p. 81.

In the case of plane sets, a rectangular area whose sides are parallel to the coordinate axes, plays the same part as a linear interval in the case of linear sets. A set contained in such a rectangle is said to be ***bounded***.

Corresponding to the fundamental principle that a series of intervals, each of which contains the subsequent ones, has one point interior to all the intervals, provided that the lengths of the intervals converge to zero, we have the principle that the points interior to a set $\delta_1, \delta_2, \ldots \delta_n, \ldots$ of rectangles each of which contains the next, consist of a single point or of a linear interval, according as both, or only one, of the pairs of sides of the rectangles have the limit zero, when n is indefinitely increased.

The theorem, that every infinite bounded plane set has at least one limiting point, is then proved by dividing the rectangle, in which the set is contained, into a finite number of parts by means of lines parallel to the axes. At least one of the resulting rectangles must contain an infinite number of points of the set either in its interior or on its boundary; choosing such a rectangle, we proceed to divide it as before into a finite number of parts, and continually apply the same argument; in all these rectangles, there is at least one point which must be a limiting point of the given set, since we may choose the mode of subdivision so that both pairs of sides of the rectangles have their limit zero. In any rectangular area whatever, which has a limiting point P of the set in its interior, there are an infinite number of points of the set.

A plane set is everywhere-dense when points of the set lie within every rectangle, with sides parallel to the axes, which can be drawn in that rectangle in which the set lies.

A plane set is non-dense when in every such rectangle another can be found which contains no points of the set.

The definition of the successive derivatives of a plane set, and the proof that all these derivatives are closed sets, is on exactly the same lines as in the case of linear sets.

101. The frontier of a set of points G in plane space or space of any number of dimensions, being defined, as in § 84, to be the set of points each of which belongs to one of the sets G, $C(G)$, and is a limiting point of the other set, it will be shewn that*:—

If the complementary set $C(G)$ exists, then the frontier of G and $C(G)$ always exists, and is a closed set.

Let P be any point of G, and P' a point of the complementary set $C(G)$, and consider those points of G which are on the straight segment PP', *i.e.* those points of which the coordinates are $\frac{x+kx'}{1+k}, \frac{y+ky'}{1+k}$, where (x, y) and

* Jordan, *Cours d'Analyse*, vol. I, p. 20.

(x', y') are the coordinates of P and P' respectively, and k denotes a positive number (including zero). The linear set of points of G on PP' has, in accordance with the theorem of § 46, an upper boundary Q. This point Q which may coincide with P, is a point of the frontier of G and $C(G)$; for if Q is a point of G it is also a limiting point of $C(G)$, and if it is a point of $C(G)$, it is a limiting point of G. Therefore, if $C(G)$ exists, there is always a frontier of G and $C(G)$. Again let $Q_1, Q_2, \ldots Q_n \ldots$ be an infinite set of points of the frontier; this set has at least one limiting point Q. Such a point Q is itself a point of the frontier; for, in the set $\{Q_n\}$, there is an infinite number of points all of which belong to G, or all to $C(G)$, of which Q is the limiting point. If these points all belong to G' and to $C(G)$, then Q belongs to G' and to $\{C(G)\}'$; if they belong to G and to $\{C(G)\}'$, then Q belongs to G' and to $\{C(G)\}'$. In either case Q is a point of the frontier; and thus, since every limiting point of the frontier belongs to it, the frontier is a closed set.

If all points of the plane belong to the frontier of G and $C(G)$, then G has no interior points. If every point of $C(G)$ belongs to the frontier, then there are no points exterior to G.

102. If (x, y) and (x', y') are two points P, P', then the positive number $\{(x-x')^2 + (y-y')^2\}^{\frac{1}{2}}$ is said to measure the *distance** of P from P'.

If P is a point of a set G_1, and P' a point of another set G_2, then the distance PP' has either a lower limit or a lower extreme value, for all pairs of points of the sets G_1, G_2. In case this lower limit, or lower extreme, is a positive number $\Delta\ (>0)$, the sets G_1 and G_2 are said to be *detached* from one another.

If two bounded and closed sets G_1, G_2 are detached from one another, they contain at least one pair of points P, P' such that their distance from one another is measured by Δ.

For let $\epsilon_1, \epsilon_2, \ldots \epsilon_n, \ldots$ be a sequence of decreasing positive numbers converging to the limit zero. A pair of points P_1, P_1' of G_1, G_2 can be determined, such that $P_1P_1'^2 < \Delta^2 + \epsilon_1$; again a pair P_2, P_2' can be determined, such that $P_2P_2'^2 < \Delta^2 + \epsilon_2$, and, in general, a pair P_n, P_n' of points can be determined, for which $P_nP_n'^2 < \Delta^2 + \epsilon_n$. If (x_n, y_n) and (x_n', y_n') are the coordinates of P_n, P_n', the coordinates (x_n, y_n, x_n', y_n') determine a point p_n in the four-dimensional continuum. The set of points $p_1, p_2, \ldots p_n, \ldots$ has at least one limiting point (x, y, x', y'); let P, P' denote the two points (x, y), (x', y') in the two-dimensional domain. It will be shewn that P, P' belong to G_1, G_2 respectively, and that PP' is measured by Δ. A number m can be found such that $x - x_n$, $y - y_n$, $x' - x_n'$, $y' - y_n'$ are all numerically less than an arbitrarily chosen positive number η, provided $n \geqq m$; it follows that P is

* Instead of the distance so defined, Jordan employs, in this connection, the "écart," defined as $|x-x'|+|y-y'|$.

a limiting point of the set $P_1, P_2, \dots P_n, \dots$, and that P' is a limiting point of the set $P_1', P_2', \dots P_n', \dots$. Since these sets belong to the closed sets G_1, G_2 respectively, it follows that P belongs to G_1, and P_2 to G_2. We have, further,

$$|x-x'| \leqq |x-x_n| + |x_n - x_n'| + |x_n' - x'| \leqq 2\eta + |x_n - x_n'|,$$

and similarly, $|y-y'| \leqq 2\eta + |y_n - y_n'|$, for $n \geqq m$. From these inequalities, we see that $(x-x')^2 + (y-y')^2 < 8\eta^2 + A\eta + P_nP_n'^2$, where A is some fixed number; hence $PP'^2 < 8\eta^2 + A\eta + \epsilon_n + \Delta^2$, and since η, ϵ_n are both arbitrarily small, it follows that $PP'^2 \leqq \Delta^2$; and thus PP', which is certainly not less than Δ, must be equal to Δ. The theorem has thus been established.

103. *A bounded and closed set of points is said to be connex or single-sheeted* (d'un seul tenant), *when it cannot be decomposed into two or more detached closed sets.*

If P, P' are any two points of a connex closed set G, then if ϵ is any positive number whatever, points $p_1, p_2, p_3, \dots p_n$ can be determined, all of which belong to the set, and are such that the distances $Pp_1, p_1p_2, p_2p_3, \dots p_nP'$ are all $\leqq \epsilon$; and conversely, if this condition is satisfied, then G is connex.

The condition stated in the theorem is sufficient to ensure the connexity of the set G. For if G can be divided into two separated closed sets G_1, G_2, such that Δ is the lower limit, or the lower extreme, of the distances of pairs of points of G_1, G_2, we may choose ϵ to be $<\Delta$. If P is a point of G_1, and p_1 is a point such that $Pp_1 < \epsilon$, the point p_1 belongs to G_1; again if p_2 is a point such that $p_1p_2 < \epsilon$, p_2 also belongs to G_1, and so on. Since p_n belongs to G_1, whatever finite value n may have, it is impossible that $p_nP' \leqq \epsilon$, because $p_nP' \geqq \Delta$. Again the condition is a necessary one. For let us suppose that, for some value of ϵ, the condition is not satisfied for every pair of points. If P be a point belonging to such a pair, the set G may be divided into two parts G_1 and G_2, where G_1 is such that, for each point P' belonging to it, a definite set of points of G_1, viz. $p_1, p_2, \dots p_n$, exists such that $Pp_1, p_1p_2, \dots p_nP'$ are all $\leqq \epsilon$, and G_2 is such that for each point of it this condition is not satisfied. The two sets G_1, G_2 are closed, and are such that the lower limit, or the lower extreme, of the distance between pairs of points in them is $>\epsilon$. For if p is a limiting point of G_1, it belongs either to G_1 or to G_2; and since there are points p_n of G_1, such that $pp_n < \epsilon$, the point p clearly belongs to G_1; therefore G_1 is a closed set. Again if q is a limiting point of G_2, it cannot belong to G_1; for a point P' of G_2 can be found such that $qP' < \epsilon$, hence if q' belonged to G_1, so also would P'. It is clear that no pair of points of G_1, G_2 can exist, of which the distance is $\leqq \epsilon$, hence for these sets $\Delta > \epsilon$. It has thus been shewn that, if for any ϵ the condition is not satisfied, G can be divided into two detached closed sets, and it is therefore not connex.

A connex closed set, which does not consist of a single point, is a perfect set.

...nt of the set could be considered as a set detached ...nsists of all the remaining points, and hence, if such an ...sted, the set could not be connex.

...be observed that a connex closed one-dimensional set can only ... a single interval.

...heory of plane sets and of sets of three or more dimensions is of ...portance in relation to its application to the *Analysis Situs*. Jordan*, ...g given an arithmetical definition of a simple closed curve, has esta-...lished the fundamental theorem that such a curve divides the plane into two parts, respectively external and internal to the curve. The subject has been further developed by Schönflies†, from the point of view of the theory of sets of points.

104. The mode in which a non-dense plane set is determined by means of areas, free in their interiors from points of the set‡, is not in all respects similar to the mode in which a non-dense linear set is determined by means of the complementary intervals. In the latter case each point P which does not belong to the set is enclosed in an interval which contains no points of the set, and this interval has a maximum length in both directions from the point P, the end-points of such maximum interval δ being points of the closed and non-dense linear set, and this maximum interval is identical with δ, for all points P interior to δ. But in the case of a plane set, if we confine ourselves to areas of given shape, such as rectangles, and these take the place of the linear intervals δ, it is not the case that a closed set is defined as the set of boundary points, together with their limits, of a unique system of such rectangles.

If P be a point which does not belong to a given non-dense plane closed set, and if we draw through P a straight line parallel to the line whose equation is $y = mx$, then those points of the given set which lie on this straight line are easily seen to form a closed set, and the point P must be interior to a complementary interval $\Delta_m(P)$, of this closed set. If on one side of P there are, in this straight line through P, no points of the given set, then on this side the extremity of the interval $\Delta_m(P)$ may be regarded as the point in which the straight line intersects a side of the rectangle in which the plane set is contained. The interval $\Delta_m(P)$ exists for every value of m, and the extremities of the intervals $\Delta_m(P)$, for a fixed P, are in general points of the plane set; the region of plane space A_P, in which these intervals $\Delta_m(P)$ lie, is free in its interior from points of the plane set; and such a region is the true analogue, for plane sets, of the complementary interval of a

* See the *Cours d'Analyse*, vol. I, pp. 90—100.

† *Göttinger Nachrichten*, 1899, also *Math. Annalen*, vols. LVIII and LIX. The subject has also been treated by Veblen, *Trans. of the American Math. Soc.* vol. VI.

‡ See Schönflies, *Göttinger Nachrichten*, 1899, p. 282, also *Bericht über die Mengenlehre*, p. 81.

linear set. The plane closed set consists of points on the boundaries of a system of such regions A_P which do not overlap, and of the limiting points of these points on the boundaries; and every point which does not belong to the set is interior to one of the regions A_P. In this sense there is for each point P of the complementary set a single region A_P which is the maximum free region containing P in its interior; and all points interior to A_P have their maximum free regions identical with A_P.

If, however, we work only with rectangular areas, which are usually the most convenient in view of applications of the theory to the theory of integration, there exists in general no rectangular area corresponding to a point P which has analogous properties to the region A_P. If we describe a square of sides 2ρ parallel to the axes of coordinates and with its centre at P, then, for any point P of the complementary set, when ρ is small enough there are no points of the given set interior to or on the boundary of the square; and ρ may be increased until one of the sides of the square contains a point of the set, or is coincident with a side of the rectangle in which the whole set is contained. When either of these things happens, we may keep this particular side fixed in position, letting the other three increase their distances from P by the same amounts; if a corner of the square comes to be a point of the set, then both the sides intersecting at that corner are kept fixed; the square now becomes a rectangle, and ultimately another side will either contain a point of the set, or will fall on a boundary of the space in which the set exists. Proceed in this way until we have a rectangle such that each of its sides contains one or more points of the plane set, or else falls upon a boundary of the domain of the set; we have then a definite rectangle corresponding to the point P. But if we take a point Q inside this rectangle, and construct the corresponding rectangle for Q, this need not coincide with the rectangle constructed for P; because a side of the rectangles, drawn with Q as centre, may come into a fixed position, by meeting a point of the given set, before it has reached the final position of the corresponding side of the rectangles constructed for P; and the maximum free rectangle for a point P, does not then, in general coincide with the maximum free rectangles for points inside the first.

105. It is however possible, for a given closed non-dense plane set G, to construct an enumerable set of rectangles which is everywhere-dense, and such that every point of G lies on the boundary of a rectangle, or is a limiting point of points which lie on the boundaries of such rectangles. Let us denote by S the rectangle in which the whole set G lies, and let δ be the rectangle constructed as above for a point P of the set G. Produce the sides of δ, when necessary, until they cut the sides of S, thus dividing S into at most nine different rectangles, of which one is δ, and the others may be denoted by S_r, where $r = 1, 2, \dots 8$. In each rectangle S_r take any point P_r

which does not belong to G, and construct for P_r the maximum free rectangle δ_r as before; let the sides of δ_r be produced when necessary until they meet the sides of S_r, then S_r is divided into at most nine rectangles, which consist of δ_r and at most eight rectangles S_{rs} when $s = 1, 2, \dots 8$.

Proceeding in this manner we obtain a set of rectangles

$$S,\ S_r,\ S_{rs},\ S_{rst}\ \dots,$$

and in them a set of rectangles $\delta, \delta_r, \delta_{rs}, \delta_{rst} \dots$, each of which contains no points of G in its interior, each of the numbers $r, s, t, \dots$ being one of the digits $1, 2, 3, \dots 8$. If p be a point of G which is not on a boundary of any rectangle δ_n, it must be in the interior of each of an unending series of rectangles $S_r, S_{rs}, S_{rst}, \dots$, where $r, s, t, \dots$ have definite values; and this set of rectangles must converge either (1) to a point in the interior of all of them, or (2) to a linear interval, or (3) to a definite rectangle S_ω in the interior of all of them. In case (1), the point to which the rectangles converge is a limiting point of those points of G which lie on the boundaries of the definite sequence of rectangles $\delta_r, \delta_{rs}, \delta_{rst}, \dots$. In case (2), there must be, on the limiting linear interval, at least one point which is a limiting point of G: for, if not, the whole interval could be enclosed in a rectangle which contains no points of G; and this is impossible. In case (3), we start with the rectangle S_ω, and take a point P_ω not belonging to G inside it, construct the maximum free rectangle δ_ω, produce its sides as before to meet those of S_ω and proceed as before to construct $S_{\omega rst} \dots$ and $\delta_{\omega rst} \dots$. This process can be continued until an index is reached which may be any number of the second class, but the point p must be reached before some definite number of the second class appears as index; this following from the fact that the number of non-overlapping regions which are contained in a given space must be enumerable. Thus the point p is reached after an enumerable set of steps of the process.

It has therefore been shewn that:—

If G is a non-dense closed plane set of points, an everywhere-dense enumerable set of rectangles can be determined, such that every point of G is on a boundary of one or more of the rectangles, or is a limiting point of such points, or lies in a linear interval which is the limit of a sequence of the rectangles.

In case the set G is perfect, the rectangles of the set must either not abut on one another, or every common side must contain either no points of G, or else a perfect set of points of G.

106. That a perfect plane set G has the power of the continuum* may be proved by projecting the set on a straight line which we may take to be a side of the rectangle in which the set is contained. The set of points

* See Bendixson, *Bib. Svensk. Vet. Handl.* vol. IX (1884), where the first proof of this theorem was given.

which are the projections of points of G is a closed set. For, if P be one of the limiting points of the set of projects, let pp' be an arbitrarily small neighbourhood of P, of which P is the centre; draw straight lines PQ, pq, $p'q'$ perpendicular to pp' to the side qq' of the containing rectangle. Then in the rectangle $pqq'p'$ there are an infinite number of points of G; and if we divide this rectangle into $(2n+1)^2$ equal parts by means of straight lines parallel to pp' and to PQ, then in one of these parts at least there are an infinite number of points of G in the interior or on one of the boundaries parallel to pp'. Also one such rectangular part, at least, exists with its centre on PQ; for otherwise P could not be a limiting point of the projection of G. Divide this rectangle into $(2n+1)^2$ equal parts as before, then in one of these at least with its centre on PQ, there must be an infinite number of points of G; proceeding in this manner we shew that there is one point at least on PQ which is a limiting point of G, and this point therefore belongs to G; thus the projection of G is a closed set. An isolated point P of the projected set must be such that there is a perfect component of P on the straight line PQ. If the projected set is perfect, then it has the power c of the continuum; and if it contains isolated points these must be the projections of perfect linear components of G; therefore in either case G has the power of the continuum.

It is clear that this method can be extended to the case of a set in any number of dimensions; and we shew that the power of an n-dimensional perfect set is c, if that of an $n-1$ dimensional perfect set is c.

107. The content of a closed plane set may be defined in a manner strictly analogous to Harnack's definition of the content of a linear closed set. If the rectangle in which the set is contained be divided into rectangular portions, by drawing a finite number of straight lines parallel to the sides of the rectangle, and the sum of the areas of those rectangular portions be taken which contain in their interiors, or on their boundaries, points of the closed set, then the content of the set is the limit of the sum when the number of the rectangular portions is increased indefinitely in such a manner that the greatest of the sides of all the rectangles has the limit zero. That the content so defined has a definite value independent of the mode in which the successive subdivisions of the original rectangular area are carried out, provided only that the greatest of all the sides of the rectangular areas diminishes indefinitely as the number of the rectangles is increased indefinitely, may be proved in precisely the same manner as in the case of linear sets.

For plane sets in general, the exterior and interior measure may be defined as in the case of linear sets.

The exterior measure $m_e(G)$ of a set G is the lower limit of the sum of the areas of a finite, or indefinitely great, number of rectangles which enclose all *the points of G in their interiors, when every possible such system of rectangles is taken account of. The interior measure $m_i(G)$ is the excess of the area of the rectangle in which G is contained over the exterior measure of the set complementary to G; or we may take the equivalent definition, that the interior measure is the upper limit of the contents of the closed components of G.*

A plane set is measurable when the exterior and interior measures have identical values.

All the theorems which have been proved in §§ 81–84 relating to the measures of linear sets hold also for plane sets, and for sets in any number of dimensions.

In the case of a set of points on a straight line, the content, or the measure of the set, considered as a set in two dimensions, is always zero, whatever value the content, or the measure of the set, considered as a linear set, may have. The latter may be spoken of as the linear content, or the linear measure of the set.

108. If G be any closed set of points in the rectangle $ABCD$, and through the points P of AB straight lines PP' are drawn perpendicular to AB, and

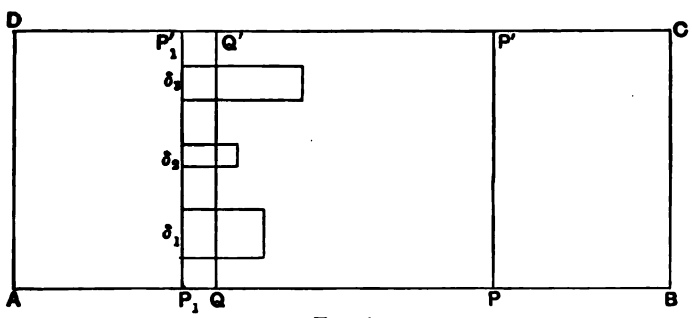


Fig. 4.

if $f(P)$ denote the linear content of the linear component of G which is on the straight line PP', then *the set of points P on AB, which is such that $f(P) \geqq \sigma$, is a closed set, σ denoting any positive number.*

Let P_1 be a limiting point of the set; and if possible, let the linear content of that component of G which is on P_1P_1' be $< \sigma$; we can then find a finite number of intervals $\delta_1, \delta_2, \ldots \delta_r$ on P_1P_1' whose sum is $> AD - \sigma$, and which are free in their interiors and at their ends from points of G. On each of these intervals δ we can describe a rectangle which contains no points of G in its interior or on its boundaries: this may be done on either side of P_1P_1'; for each point of δ can be enclosed in a rectangle free from points of G, and by the Heine-Borel theorem, a finite number of these rectangles, enclosing all the points of δ, exists. Take a point Q belonging to the set of points for which $f(Q) \geqq \sigma$, and let P_1Q be less than the breadth of all the rectangles described on the intervals $\{\delta\}$ on one side of P_1P_1'. On QQ' there is a finite number of intervals free from points of G, whose sum is $> AD - \sigma$, by the assumption as to P_1P_1'; hence the linear content of the component of G which is on QQ' must be $< \sigma$, which is contrary to the hypothesis. It follows that $f(P_1) \geqq \sigma$; hence the set of points on AB is closed.

It will now be shewn that, *for a closed set of points G, if for every position of P on AB, the linear content of the component of G upon PP' is $< \sigma$, then the content of G is $< \sigma . AB$.*

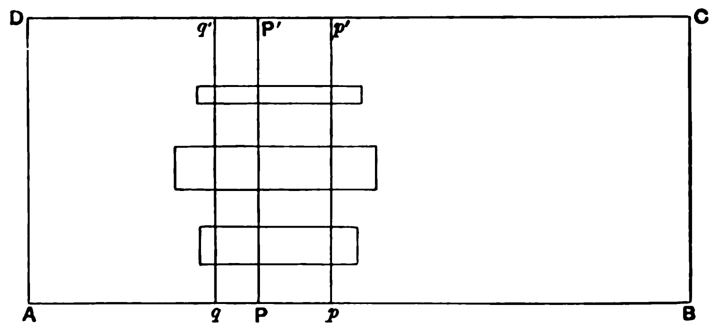


FIG. 5.

Taking any point P of AB, on PP' a finite number of intervals, whose sum is $> AD - \sigma$, can be found which are free from points of G; and on each of these intervals a rectangle can be drawn on each side of PP' containing no points of G in its interior or on its boundary. We can now draw two straight lines pp', qq', one on each side of P, so that each of them passes through the interiors of all the rectangles so described. We have now found an interval pq containing P, such that in $pqq'p'$ there is an area

$$> pq(AD - \sigma)$$

free from points of G. Corresponding to each point P of AB such an interval pq can be found; and, in accordance with the Heine-Borel theorem, a finite number of these intervals can be selected such that every point of AB is in the interior of one at least of them. The end-points of these intervals divide AB into a finite number of parts such that, above any one part of length α, there is an area $> \alpha(AD - \sigma)$ free from points of G; and hence there is altogether an area $> AB(AD - \sigma)$ free from points of G. It follows therefore that the content of G is $< AB . \sigma$.

We shall now establish the following theorem, which is of importance in the theory of double integration:—

If G be a closed set, and if the linear content of the set of points P on AB for which the linear content of that component of G, which lies on PP', is $\geqq \sigma$, have the value zero for every positive value of σ, then the set G is of zero content.

The points on AB, for which $f(P) \geqq \sigma$, can be enclosed in a finite number of intervals whose sum is $< \epsilon$, where ϵ is an arbitrarily small number; and in each of the remaining parts of AB, the value of $f(P)$ is $< \sigma$; hence by the foregoing theorem the content of G is $< \sigma(AB - \epsilon) + \epsilon . AD$; and since this holds for arbitrarily small values of σ and ϵ, it follows that the content of G must be zero.

Conversely, it may be shewn that:—

If G be a closed set of zero plane content, the set of points P on AB, for which the linear content of the component of G on PP' is $\geqq \sigma$, has, for every positive value of σ, content zero.

Let I denote the linear content of the set (P) for which $f(P) \geqq \sigma$; divide AB into n equal parts, and AD also into n equal parts, and through the end-points of these parts draw straight lines dividing the rectangle into equal parts each of area $\frac{1}{n^2} . AB . AC$. Then the sum of those parts of AB which contain points of the set (P), is always greater than I; and in each such part there is at least one point P, such that the sum of the parts of PP' which contain points of G is $> \sigma$. It follows that the sum of those rectangular portions which contain points of G is $> \sigma I$, however great n may be; and hence that the content of G is $\geqq \sigma I$. Therefore it follows that G cannot have zero content unless I is zero.

EXAMPLES.

1. Let a set of points (x, y) in the rectangle for which $0 \leqq x \leqq 1$, $0 \leqq y \leqq 1$, be defined as follows*:—The numbers x, y are expressed in the dyad scale, and only those values of x and y are taken which are expressed by terminating radix-fractions, the number of digits being the same for x as for y. If x' denotes a terminating radix-fraction, there are only a finite number of points (x', y') of the set on the straight line $x = x'$; similarly if y' denotes a terminating radix-fraction, there are only a finite number of points of the set on the straight line $y = y'$. The two-dimensional set is however everywhere-dense; for, considering a straight line $y = x + a$, where a is a positive or negative radix-fraction with a finite number of digits, we see that, corresponding to any number x expressed by a finite number of digits greater than the number of digits by which a is expressed, there is a point (x, y) on the straight line belonging to the set. The component of the set on the straight line $y = x + a$, being everywhere-dense, and the values of a being everywhere-dense in the interval $(-1, 1)$, it follows that the set is everywhere-dense in the rectangle.

This example shews that an everywhere-dense two-dimensional set may be linearly non-dense on each straight line belonging to two parallel sets. It also shews that a two-dimensional set may exist which is extended, but is unextended on straight lines belonging to either of two parallel sets.

2. Let a cross† formed by two pairs of straight lines parallel to the pairs of sides of a square be constructed, and so that the remainder of the square consists of four equal

* Pringsheim, *Sitzungsberichte d. Münchener. Akad.* vol. XXIX, p. 48.

† Veltmann, *Schlömilch's Zeitsch.* vol. XXVII, pp. 178, 314.

squares at the corners. Let the interior points of the cross be removed from the square, and then let a similar cross be removed from each of the remaining four squares. Proceeding in this manner, let the crosses be so chosen that the area of each square after the mth stage of the process is $ab^{-\frac{p}{2^m}}$ times the area of each square after the preceding stage. The sum of the areas of the squares which remain after the mth stage is

$$(4a)^m b^{-p\left(\frac{1}{2^m}+\frac{1}{2^{m-1}}+\dots+\frac{1}{2}\right)}Q=(4a)^m b^{-p\left(1-\frac{1}{2^m}\right)}Q,$$

where Q is the area of the original square. A non-dense closed set of points is defined as the points which remain when this process is carried on indefinitely. The limit of the sum of the crosses is that of

$$\left[1-(4a)^m b^{-p\left(1-\frac{1}{2^m}\right)}\right]Q;$$

and this is Q or $< Q$, according as $a \leqq \frac{1}{4}$; it follows that the closed set has zero content if $a < \frac{1}{4}$, but if $a = \frac{1}{4}$, the content is $b^{-p}Q$.

SETS OF SEQUENCES OF INTEGERS.

109. A theory of sets of sequences of integers, of which the formal character is similar to the theory of sets of points in any number of dimensions, has been developed by Baire*, with a view to application to the Theory of Functions.

A *group* of integers $(\alpha_1, \alpha_2, \dots \alpha_p)$, of order p, consists of a system of p positive integers arranged in a definite order.

The group $(\alpha_1, \alpha_2, \dots \alpha_p)$, of order p, is said to be contained in each of the groups (α_1), (α_1, α_2), $(\alpha_1, \alpha_2, \alpha_3), \dots (\alpha_1, \alpha_2, \dots \alpha_{p-1})$ of orders $1, 2, 3, \dots p-1$, respectively.

A *sequence* of integers $(\alpha_1, \alpha_2, \dots \alpha_p, \dots)$ consists of an infinite number of integers, defined in any manner, and arranged in an order similar to the sequence $1, 2, 3, \dots$. This sequence is said to be contained in each of the groups (α_1), $(\alpha_1, \alpha_2), \dots (\alpha_1, \alpha_2, \dots \alpha_p), \dots$. Let P be a set of such sequences of integers, and let A be any other sequence of integers; then if, for every n, there are sequences in P, other than A itself, which are contained in the same group of order n as A itself is contained in, *i.e.* sequences having their first n integers the same as the first n integers in A, then the sequence A is said to be a *limit* of the set of sequences P. The sequence A may or may not itself belong to P.

The set P is said to be *closed*, in case all its limits belong to it. The set is said to be *perfect* when it is closed, and also every sequence in the set is a limit of the set.

A set E of groups of integers is said to be *complete* if, when g is any group of order p belonging to E, the groups of orders $1, 2, 3, \dots p-1$, which contain g, also belong to E.

* *Comptes Rendus*, vol. CXXIX, 1899, p. 946.

A complete set E of groups of integers is said to be *closed*, if every group g belonging to E contains at least one group of higher order than itself, which is also contained in E.

Having given a complete set of groups E, a sequence A may exist such that all the groups containing A belong to E. The set F of all sequences such as A, is said to be *determined* by the set of groups E. The set F, if it exists, is closed.

Every closed set of groups E determines a closed set of sequences F, and conversely, every closed set of sequences F is determined by a unique closed set of groups E. In case F is perfect, E is also said to be perfect. In order that E may be perfect, it is necessary and sufficient that every group belonging to E should contain at least two groups of one and the same order superior to its own order, and belonging to E.

If P is a set of sequences, then the set P' of those sequences which are limits of the set P is said to be the derived set of P, and may be denoted by P'. The derived set P' is closed.

The successive derivatives $P'', P''', \ldots P^{(\omega)}, \ldots P^{(\alpha)}$, of finite or transfinite orders, are then defined as in the theory of sets of points. If P is a closed set of sequences, there exists a number α of the first or the second class, such that $P^{(\alpha)} = P^{(\alpha+1)}$. Unless P is an enumerable set, it can be resolved into the sum of an enumerable set and a perfect set.

Let us consider a perfect set of groups E determining a perfect set of sequences F. A set P of sequences all belonging to F is said to be *non-dense* in F or in E, provided that every group of E contains at least one group of E which contains no sequence of P.

A set of sequences P all belonging to F is said to be of the *first category*, relative to F, if there exists an enumerable sequence of sets $P_1, P_2, \ldots P_n, \ldots$, each of which is non-dense in F, and such that each sequence of P is part of one at least of the sets $P_1, P_2, \ldots P_n, \ldots$. The set obtained by removing the set P from F is said to be of the *second category* relative to F. The same generic distinction between sets of the first and of the second category holds, as in the theory of sets of points.

CHAPTER III.

TRANSFINITE NUMBERS AND ORDER-TYPES.

110. A PRELIMINARY account has been given in Chapter II, of the theory of transfinite ordinal and cardinal numbers; it was shewn that the introduction of such numbers was suggested by the exigencies of the theory of linear sets of points, and that, in particular, the necessity for the use of transfinite ordinal numbers arises whenever a convergent sequence of points is transcended by adjoining to the points of the sequence their limiting point and any further points which it may be desirable to regard as belonging to the same set as the points of the sequence. The fundamental discovery of G. Cantor, that the rational points of an interval form an enumerable set, whereas the set of points of the continuum is unenumerable, by establishing the existence of a distinction between the characters of two infinite sets, suggests the development of a general theory of cardinal numbers of infinite aggregates. The procedure we adopted, of introducing the fundamental notions of transfinite ordinal and cardinal numbers in connection with the theory of sets of points, is in accord with the historical order in which the whole theory of transfinite numbers and order-types was developed. The account of the theory of transfinite numbers given in Chapter II, is in general agreement with Cantor's earlier presentation* of his ideas; his later† and more abstract treatment of the subject is the one upon which the account given in the present Chapter is founded.

In order that the reader may be put into a position to form his own conclusions as to the validity of a scheme which must be regarded as still, to some extent at least, in the controversial stage, it has been thought best to postpone any discussion of the difficulties of the theory, until after the conclusion of the detailed account of the theory in its constructive aspect.

* See his "Grundlagen einer allgemeinen Mannigfaltigkeitslehre," Leipzig, 1883, or *Math. Annalen*, vol. XXI; see also *Zeitschrift für Phil. und phil. Kritik*, vols. LXXXVIII, XCI and XCII. Cantor's ideas were foreshadowed in a paradoxical form by Bolzano in his "Paradoxien des Unendlichen," Leipzig, 1851; and although infinite numbers had been discussed by earlier writers, Bolzano is the only real predecessor of Cantor in this department of thought.

† This is contained in the two articles "Beiträge zur Begründung der transfiniten Mengenlehre," in the *Math. Annalen*, vol. XLVI (1895), and vol. XLIX (1897).

In the last part of the Chapter, some critical remarks upon the logical basis of the theory will be made; these must necessarily be of an incomplete character, partly from considerations of space, and also because any complete criticism of such a scheme as Cantor's theory of transfinite numbers would involve the consideration of questions of an epistemological character which for obvious reasons cannot be adequately dealt with in a work of a professedly mathematical complexion. Objections which may be urged against some parts of the theory, will however be fully stated. Some consideration will also be given to the question, whether, and how far, the theory is indispensable as a logical basis of continuous Analysis.

THE CARDINAL NUMBER OF AN AGGREGATE.

111. *A collection* of definite distinct objects which is regarded as a single whole is called an aggregate.*

An aggregate may be denoted symbolically by a large letter M, the elements of the aggregate by small letters m; and the constitution of the aggregate may be denoted by the equation $M = \{m\}$.

The consideration of questions which arise in connection with this definition, as to the mode in which the objects of the aggregate must be specified, in order that the aggregate may be adequately defined, and as regards the conditions, if any, which must be satisfied in order that a collection may be regarded as a whole, or aggregate, of such a character that it can be an object of mathematical thought, will be postponed. For the present, it is sufficient to remark, that an adequate definition of any particular aggregate, which is not necessarily finite, must contain, as a minimum, a set of rules or specifications by means of which it is *theoretically* determinate, in respect of any object whatever, whether such object does or does not belong to the aggregate. The set of prime numbers, for example, is regarded as an aggregate, although when a particular number is presented to us, we may be *practically* unable to decide whether that number is prime or not. In this case however, a finite number of processes will suffice to decide the question. If however, we take the case of the algebraical numbers, the state of things is different; for we are not in possession of any general method which enables us to decide whether a given number is algebraic or not. Nevertheless, the question being regarded as having a definite answer, the algebraical numbers are regarded as forming an aggregate, in the sense here employed.

* This definition is given by Cantor, *Math. Annalen*, vol. XLVI, p. 481, as follows:—"Unter einer 'Menge' verstehen wir jede Zusammenfassung M von bestimmten wohl unterschiedenen Objecten m unserer Anschauung oder unseres Denkens (welche die 'Elemente' von M genannt werden) zu einem Ganzen."

An aggregate does not depend, for its validity as a mathematical entity, upon the possibility of producing all its members, successively or otherwise, but upon the sufficiency of the rules by which its elements are to be distinguished, as belonging to it, in that particular kind of objects to which they belong; that is, upon the sufficiency, in this direction, of its definition of membership.

Two aggregates M, N are said to be *equivalent* to one another when they are such that a law of correspondence can be established between the elements of one aggregate and those of the other, such that to each element of one of the aggregates, there corresponds one and only one element of the other aggregate.

This relation of equivalence between two aggregates M and N, may be expressed symbolically by $M \sim N$, or $N \sim M$.

It is clear that, if each of two aggregates is equivalent to a third, the two aggregates are equivalent to one another.

Aggregates which are equivalent to one another are said to have the same power or cardinal number.

A cardinal number is accordingly characteristic of a class of equivalent aggregates.

The question whether two defined aggregates have or have not the same cardinal number, is thus equivalent to the question whether it is, or is not, possible to establish a systematic (1, 1) correspondence between the elements of the two aggregates, in accordance with the above definition of equivalence.

A particular aggregate can ordinarily be shewn to be equivalent to itself. The law of correspondence between an element and another element which can be set up, is in general of a character which admits of a certain arbitrariness. The cardinal number is accordingly regarded as independent of the notion of order in the aggregate.

The power or cardinal number of an aggregate M has been defined by Cantor as the concept which is obtained by abstraction when the nature of the elements of M, and the order in which they are given, are entirely disregarded.

Cantor regards the fact, that equivalent aggregates have the same cardinal number, as a deduction from this definition. Some critical remarks upon the definition of the cardinal number of an aggregate will be made in § 155.

The cardinal number of M is a characteristic of M which may be denoted by $\overline{\overline{M}}$, to indicate that both the order of the elements, and their precise individual nature, are irrelevant as regards the cardinal number.

The relation of equivalence $M \sim N$, between two aggregates, implies the equality $\overline{\overline{M}} = \overline{\overline{N}}$; and this equation expresses the necessary and sufficient condition for the equivalence of M and N.

Since Cantor regards the cardinal number of M as independent of the precise nature of the elements of M, we may in accordance with this view, substitute for each element the number unity. We have thus a new aggregate which is a collection of elements each of which is the number 1, and is equivalent to M; and this new aggregate is regarded by Cantor as a symbolical representation of the cardinal number $\overline{\overline{M}}$.

THE RELATIVE ORDER OF CARDINAL NUMBERS.

112. Every aggregate M_1, which is such that all its elements are also elements of M, is called a *part* or *sub-aggregate* of M.

If M_2 is a part of M_1, and M_1 is a part of M, then M_2 is a part of M.

A finite aggregate cannot be equivalent to any of its sub-aggregates; but, as will be seen in detail further on, an infinite aggregate always possesses sub-aggregates which are equivalent to itself. This is the characteristic distinction between finite and infinite aggregates, and has in fact been employed by Dedekind and others to define an infinite aggregate, as one which is equivalent to one of its parts.

If two aggregates M, N with the cardinal numbers $\alpha \equiv \overline{\overline{M}}$, $\beta \equiv \overline{\overline{N}}$, are such that, (1) there exists no part of M which is equivalent to N, and (2) there exists a part N_1 of N which is equivalent to M, it is clear that the corresponding conditions are satisfied for any two aggregates which are equivalent to M, N respectively; and thus the two conditions characterise a relation between the cardinal numbers α, β of the two aggregates. When the above conditions are satisfied we say that α is *less* than β, and that β is *greater* than α; which is expressed symbolically by $\alpha < \beta$, $\beta > \alpha$. This is the definition of inequality for two cardinal numbers, and of the relations greater and less in the purely ordinal sense in which they are here used.

The condition contained in the definition is inconsistent with the relation of equality between α and β being satisfied. For if $\alpha \equiv \beta$, then $M \sim N$, hence since $N_1 \sim M$, we have $N_1 \sim N$: therefore, since $M \sim N$, there must be a part of M, say M_1, such that $M_1 \sim M$, which would involve $M_1 \sim N$; but this is contrary to one of the conditions contained in the definition of inequality.

It is easily seen that if $\alpha < \beta$, and $\beta < \gamma$, than $\alpha < \gamma$.

113. It has been seen that the three relations $\alpha = \beta$, $\alpha < \beta$, $\beta > \alpha$ are mutually exclusive; but the question arises whether any two cardinal numbers α, β whatever must satisfy one of these relations. An affirmative answer to this question would be required before it could be maintained that all cardinal numbers can be regarded as being alike capable of having relative rank assigned to them, in a single ordered aggregate.

Two aggregates M, N of which we may denote parts by M_1, N_1, must satisfy one and only one of the following four conditions:—

(1) M, N have parts M_1, N_1, such that $M_1 \sim N$, and $N_1 \sim M$.

(2) M has a part M_1, such that $M_1 \sim N$; but no part of N exists which is equivalent to M.

(3) There is no M_1 which is equivalent to N; but there is an N_1 which is equivalent to M.

(4) There exists no M_1 equivalent to N; and also no N_1 equivalent to M.

It will be proved that, if the condition (1) is satisfied, then $\overline{\overline{M}} = \overline{\overline{N}}$. The condition (2) expresses the relation defined as $\overline{\overline{M}} > \overline{\overline{N}}$. The condition (3) expresses the relation defined as $\overline{\overline{M}} < \overline{\overline{N}}$.

It has not yet been proved that the relation (4) is an impossible one; except that, in the case of finite aggregates, it may be easily seen that it involves $\overline{\overline{M}} = \overline{\overline{N}}$. Until this point is cleared up, it cannot be maintained as an established fact that the cardinal numbers α, β of any two aggregates whatever satisfy one of the three relations $\alpha = \beta$, $\alpha > \beta$, $\alpha < \beta$.

Two aggregates which are such that their cardinal numbers α, β stand to one another in one of the relations $\alpha = \beta$, $\alpha > \beta$, or $\alpha < \beta$, may be said to be *comparable* with one another. Otherwise they are *incomparable* with one another.

THE ADDITION AND MULTIPLICATION OF CARDINAL NUMBERS.

114. If M, N are two aggregates which have no element in common, then the aggregate which has for its elements all those of M and all those of N is called the sum of the two aggregates M, N, and may be denoted by (M, N). A similar definition applies to the case of the sum of any number of aggregates no two of which have an element in common.

If M', N' are two other aggregates with no element in common, such that $M \sim M'$, $N \sim N'$, it is clear that $(M, N) \sim (M', N')$; and thus the cardinal number of (M, N) depends only on those of M and N.

If $\overline{\overline{M}} = \alpha$, $\overline{\overline{N}} = \beta$, we define the result of the operation of addition of α and β to be $\overline{\overline{(M, N)}}$.

From the independence of cardinal numbers of the order of elements, we deduce

$$\alpha + \beta = \beta + \alpha, \quad \alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma;$$

thus the operation of addition of cardinal numbers obeys the commutative and associative laws.

115. If an element m of M be associated with an element n of N, so as to form a new element (m, n), the aggregate of all possible elements which can be formed in this way is called the product of M and N, and may be denoted by $(M.N)$.

If $M \sim M'$, $N \sim N'$, it is clear that to each element (m, n) of $(M.N)$, there is a corresponding element of $(M'.N')$, hence $(M.N) \sim (M'.N')$, and thus $\overline{\overline{(M.N)}}$ depends only on $\overline{\overline{M}}$ and $\overline{\overline{N}}$.

The cardinal number of the product-aggregate $(M.N)$ is defined to be the product of the cardinal numbers of M and N.

The product of $\overline{\overline{M}}$ and $\overline{\overline{N}}$ may also be defined as the cardinal number of the aggregate which is obtained by substituting for each element of N, an aggregate which is equivalent to M.

It is seen on reflection that this definition is equivalent to the first one.

Since, as can be shewn from the definition,

$$(M.N) \sim (N.M), \quad (M.(N.R)) \sim ((M.N).R)$$

and

$$(M.(N, R)) \sim ((M.N), (M.R)),$$

we see that cardinal numbers satisfy the relations

$$\alpha.\beta = \beta.\alpha, \quad \alpha.(\beta.\gamma) = (\alpha.\beta)\gamma, \quad \alpha(\beta+\gamma) = \alpha\beta + \alpha\gamma.$$

It has thus been shewn that the multiplication of cardinal numbers obeys the commutative, associative, and distributive laws.

The definition of multiplication may be extended* to the case in which the number of factors is not necessarily finite. Let us consider a class of aggregates M, where the class contains either a finite or an infinite number of aggregates, and suppose no two of the aggregates have an element in common. Let there be chosen from each of the aggregates in the class, one element, and conceive that this is done in every possible way; we have now a new aggregate, each element of which consists of an association of elements, one from each of the aggregates of the given class. The new aggregate is said to be the product-aggregate of the given class of aggregates, and its cardinal number is defined to be the product of the cardinal numbers of all the aggregates of the given class.

CARDINAL NUMBERS AS EXPONENTS.

116. If we have two finite aggregates M, N containing x and y elements respectively, we may suppose that to each of the y elements of N, one element of M is made to correspond, so that the same element of M may be used any number of times; any particular such correspondence we call

* See Whitehead, *American Journal of Math.* vol. xxiv, where the theory of cardinal numbers is treated by the Peano-Russell symbolical method.

a *covering* (Belegung) of N by M. The total number of ways of covering N by M is x^y. To put the matter in a concrete form, the total number of ways of distributing y things among x persons, where any number of the y things may be given to one person, is x^y; any particular mode of distribution is what we have called a mode of covering the aggregate of y things by the aggregate of x persons.

The definition of covering an aggregate N by an aggregate M is immediately extensible to the case of infinite aggregates. As before, the covering denotes any system by which to each element of N is made to correspond a particular element of M, the same element of M being employed any number of times, or not at all. Denoting by N/M each particular mode of covering N by M, we thus form the new aggregate (N/M) which contains as its elements all such coverings.

It is seen at once that, if $M \sim M'$, $N \sim N'$, then $(N/M) \sim (N'/M')$. Thus the cardinal number of (N/M) depends only on the cardinal numbers of M and N.

The cardinal number of the aggregate (N/M), each element of which is a covering of N by M, and in which every possible mode of such covering occurs as an element, is denoted by the symbol α^β, where $\alpha \equiv \overline{\overline{M}}$, $\beta \equiv \overline{\overline{N}}$; thus $\alpha^\beta \equiv \overline{\overline{(N/M)}}$.

It is easy to shew that

$$((N/M).(R/M)) \sim ((N, R)/M)$$
$$((R/M).(R/N)) \sim (R/(M.N))$$
$$(R/(N/M)) \sim ((R.N)/M).$$

Hence if $\overline{\overline{M}} = \alpha$, $\overline{\overline{N}} = \beta$, $\overline{\overline{R}} = \gamma$, we see that, in accordance with the above definition of exponentials,

$$\alpha^\beta . \alpha^\gamma = \alpha^{\beta+\gamma}, \quad \alpha^\gamma . \beta^\gamma = (\alpha . \beta)^\gamma, \quad (\alpha^\beta)^\gamma = \alpha^{\beta . \gamma};$$

and thus the same laws hold as for exponents in which only finite cardinal numbers are involved.

THE SMALLEST TRANSFINITE CARDINAL NUMBER.

117. The cardinal number of the aggregate of all the finite integers 1, 2, 3, ... n, ... is called Alef-zero, and is denoted by $\aleph_0$; thus $\aleph_0 = \{\overline{\overline{n}}\}$. The number $\aleph_0$ is identical with the number which has previously been denoted by a.

If we add to $\{n\}$ a new element e, we obtain the sum-aggregate $(\{n\}, e)$, and this is equivalent to $\{n\}$, for we may make e in the first of these aggregates correspond to 1 in the second, and in general n to $n+1$; and thus $(\{n\}, e) \sim \{n\}$. From this, we obtain $\aleph_0 + 1 = \aleph_0$, a relation which differentiates $\aleph_0$ from all the finite cardinal numbers.

The cardinal number $\aleph_0$ is greater than all the finite cardinal numbers, and it is less than any other transfinite cardinal number.

Since the finite aggregate $(1, 2, 3, \ldots k)$ is a part of $\{n\}$, but no part of the finite aggregate is equivalent to $\{n\}$, by the definition of inequality we have $\aleph_0 > k$.

To prove that, if α is any transfinite number, say that of an aggregate M, which is not equivalent to $\{n\}$, then $\aleph_0 < \alpha$, we have to shew that M contains a part which is equivalent to $\{n\}$, and that there exists no part of $\{n\}$ which is equivalent to M. A first, second, third, ... nth element can be chosen from the elements of M in any manner, and this process can be continued without limit; thus M always contains a part which is equivalent to $\{n\}$. Any part of the aggregate $\{n\}$ which is not finite, consists of finite numbers chosen from $1, 2, 3, \ldots n, \ldots$; and of these there must be one which is smallest: denote it by e_1. Then the next greater can be denoted by e_2, and so on. Thus this part of $\{n\}$ is $(e_1, e_2, e_3, \ldots)$, which is equivalent to $\{n\}$; and therefore the theorem is established.

118. It has been shewn that $\aleph_0 + 1 = \aleph_0$: a similar proof would shew that $\aleph_0 + n = \aleph_0$, where n is any finite integer.

In accordance with the definition of addition, $\aleph_0 + \aleph_0$ is the cardinal number of the aggregate $(1, 3, 5, \ldots 2, 4, 6, \ldots)$, for $\aleph_0$ is the cardinal number of each of the aggregates $(1, 3, 5, \ldots)$ $(2, 4, 6, \ldots)$; hence, since the cardinal number of $(1, 3, 5, \ldots 2, 4, 6, \ldots)$ is the same as that of $\{n\}$, we have $\aleph_0 + \aleph_0 = \aleph_0$, which we may write as

$$\aleph_0 . 2 = 2 . \aleph_0 = \aleph_0.$$

From this relation, by repeated addition of $\aleph_0$ to both sides of the identity, we find

$$\aleph_0 . n = n . \aleph_0 = \aleph_0.$$

In order to express the product $\aleph_0 . \aleph_0$ we form the aggregate $\{(n, n')\}$ of which the elements (n, n') consist of every pair of finite cardinal numbers. Let $n + n' = s$, then s has the values $2, 3, 4, \ldots$; and for any fixed value of s the numbers n, n' have a definite number of sets of values. Let $s = 2$, we then have one element $(1, 1)$: let $s = 3$, we then have two elements $(1, 2)$, $(2, 1)$: for $s = 4$, we have $(1, 3)$, $(2, 2)$, $(3, 1)$, and so on. The elements of $\{(n, n')\}$ may thus be arranged in order so that the element (n, n') is at the pth place, when $p = n + \frac{(n + n' - 1)(n + n' - 2)}{2}$; thus the aggregate $\{(n, n')\}$ is equivalent to $\{p\}$, which has the cardinal number $\aleph_0$.

It has now been proved that $\aleph_0 . \aleph_0 = \aleph_0$, or $\aleph_0^2 = \aleph_0$; and from this the theorem $\aleph_0^n = \aleph_0$, follows by repeated multiplication by $\aleph_0$.

The theorems $n . \aleph_0 = \aleph_0$, $\aleph_0^2 = \aleph_0$, express in a symbolical form the results which have been proved in § 54, that a finite, or enumerably infinite, number of enumerable aggregates makes an enumerable aggregate.

THE EQUIVALENCE THEOREM.

119. The proof referred to in § 113, will now be given, that if M, N are any two aggregates such that M contains a part M_1 which is equivalent to N, and N contains a part N_1 equivalent to M, then $\overline{\overline{M}} = \overline{\overline{N}}$. This theorem, which may be called the equivalence theorem, was first proved by Schröder* and independently by Bernstein†; but the form in which the proof is here given is due to Zermelo‡.

Lemma I. If a cardinal number α remains unaltered by the addition of any one of the enumerable set of cardinal numbers $p_1, p_2, \ldots p_n, \ldots$, it remains unaltered if all these cardinal numbers p are added to it at once.

If $M, P_1, P_2, \ldots P_n, \ldots$ be aggregates of which the cardinal numbers are $\alpha, p_1, p_2, \ldots p_n, \ldots$; and such that $P_1, P_2, \ldots P_n, \ldots$ are all parts of M. We have then, $M = (P_1, M_1) = (P_2, M_2) = \ldots = (P_n, M_n) \ldots$; where $M_1, M_2, \ldots$ are all parts of M, and in virtue of the hypothesis made in the statement of the theorem,

$$\overline{\overline{M}} = \overline{\overline{M}}_1 = \overline{\overline{M}}_2 = \ldots = \overline{\overline{M}}_n \ldots.$$

We may denote the (1, 1) correspondence which can be set up (see § 111) between M and M_n, by $M_n = \phi_n M$; and this for every n. Now it is clear that this relation of correspondence is such that

$$\phi M = (\phi P_1, \phi M_1) = (\phi P_2, \phi M_2) = \ldots.$$

Hence $$M = (P_1, M_1),$$

$$M_1 = \phi_1 M = (\phi_1 P_2, \phi_1 M_2) = (P_2', M_2'),$$

where P_2', M_2' are those aggregates which correspond to P_2, M_2 respectively in the correspondence denoted by ϕ_1.

Also, with a similar notation,

$$M_2' = \phi_1\phi_2 M = (\phi_1\phi_2 P_3, \phi_1\phi_2 M_3) = (P_3', M_3'),$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,$$

$$M'_{r-1} = \phi_1\phi_2 \ldots \phi_{r-1} M = (\phi_1\phi_2 \ldots \phi_{r-1} P_r, \phi_1\phi_2 \ldots \phi_{r-1} M_r) = (P_r', M_r').$$

From these results we deduce

$$M = (P_1, P_2', P_3', \ldots P_r', M_r');$$

and no two of the parts $P_1, P_2', P_3', \ldots P_r'$ of M, have elements in common.

* See *Jahresbericht d. Deutsch. Math. Verg.* vol. v, p. 81 (1896); also *Nova Acta Leop.* vol. LXXI, p. 303 (1898).

† See Borel's *Leçons sur la théorie des fonctions*, p. 103.

‡ *Göttinger Nachrichten*, 1901, p. 34, "Ueber die Addition transfiniter Cardinalzahlen."

This process of division of M can be continued indefinitely; and we then have

$$M = (P_1, P_2', P_3', \dots M_\omega'),$$

where P_r' for every r is included, and M_ω' consists of those elements which belong to M_r' for every value of r. From this we see that

$$\alpha = p_1 + p_2 + p_3 + \dots + \alpha';$$

where α' is the cardinal number of M_ω'.

Let us now consider the special case of the lemma which arises when $p_1, p_2, \dots$ are all equal, say to p. In this case, we see that, from the hypothesis $\alpha = p + \alpha$, the result $\alpha = \aleph_0 p + \alpha'$ follows, where $\aleph_0$ denotes the cardinal number of the series of finite integers.

Now since $\aleph_0 = 2\aleph_0$, we have $\aleph_0 p + \alpha' = 2\aleph_0 p + \alpha' = \aleph_0 p + \alpha$; it has thus been shewn that if $\alpha = \alpha + p$, then $\alpha = \alpha + \aleph_0 p$.

Returning to the general case, we have

$$\alpha = \alpha + \aleph_0 p_1 = \alpha + \aleph_0 p_2 = \dots;$$

it now follows that $\alpha = \aleph_0 p_1 + \aleph_0 p_2 + \dots + \alpha''$,

where α'' is the value which α' takes when $\aleph_0 p_1, \aleph_0 p_2, \dots$ are substituted for $p_1, p_2, \dots$.

We now have

$$\alpha = 2\aleph_0 (p_1 + p_2 + \dots) + \alpha'' = \alpha + \aleph_0 (p_1 + p_2 + \dots)$$
$$= (\aleph_0 + 1)(p_1 + p_2 + \dots) + \alpha'' = \alpha + p_1 + p_2 + \dots:$$

and therefore the Lemma has been established in an extended form.

Lemma II. If the sum of two cardinal numbers p and q when added to α leaves α unaltered, then α is unaltered by the addition of either p or q.

For if $\alpha = \alpha + p + q$,

we have seen that $\alpha = \alpha + \aleph_0 (p + q)$;

hence $\alpha = \alpha + (\aleph_0 + 1) p + \aleph_0 q$,

and also $\alpha = \alpha + \aleph_0 p + (\aleph_0 + 1) q$;

from these equalities we have

$$\alpha = \alpha + p, \text{ and } \alpha = \alpha + q.$$

We are now in a position to prove the equivalence theorem. If

$$\alpha = \beta + p, \text{ and } \beta = \alpha + q,$$

we have $\alpha = \alpha + p + q$,

and hence, by Lemma II, $\alpha = \alpha + p = \alpha + q = \beta$;

therefore if M has a part equivalent to N, and N has a part equivalent to M, it follows that $\overline{\overline{M}} \equiv \alpha = \beta \equiv \overline{\overline{N}}$.

In case the condition $\alpha=\beta+p$, holds, but there is no corresponding condition $\beta=\alpha+q$, we have in accordance with the definition in § 112, $\alpha>\beta$. It follows that *the sum of two or more cardinal numbers is greater than, or equal to, any one of the cardinal numbers.*

The following theorem may be established:—

If the cardinal number α is unaltered by the addition of p, then if $\beta \geqq \alpha$, the cardinal number β is unaltered by the addition of q, where $q \leqq p$.

For let $\beta=\alpha+\gamma$, $p=q+r$; then from $\alpha=\alpha+p=\alpha+q+r$, we deduce that $\alpha=\alpha+q$. It then follows that

$$\beta=\alpha+\gamma=\alpha+q+\gamma=\alpha+r+\gamma=\beta+q=\beta+r.$$

120. A proof has been given by Cantor* that *if an aggregate exists of which the cardinal number is α, then an aggregate always exists of which the cardinal number is greater than α.*

The proof is a generalization of the second proof given in § 56, that the cardinal number c of the continuum is greater than that of the rational numbers. The proof may be put into the following form:—

Suppose $M=\{m\}$ to be an aggregate of cardinal number α; which aggregate M may be supposed to be simply ordered in any manner. In M let each element m be replaced either by A or by B, where A, B are two given objects; then M is replaced by a similar aggregate (see § 122), in which each element is either A or B. An infinity of such aggregates will be obtained differing from one another in respect of whether A or B has been put in the place of each element of M; denoting the aggregate of all such possible aggregates M_{AB}, by $\{M_{AB}\}$, it will be shewn that the cardinal number of $\{M_{AB}\}$ is greater than that of M. In the first place, it can be seen that the cardinal number of $\{M_{AB}\}$ is equal to, or greater than, that of $\{m\}$; for, taking any one element m_0 of $\{m\}$, replace it by A, and all the other elements by B; we have then an element of $\{M_{AB}\}$, and there is such an element corresponding to each element m_0 of $\{m\}$; thus those elements of $\{M_{AB}\}$, in which there is only one A, form an aggregate of cardinal number equal to that of $\{m\}$. Next, let us assume that, if possible, all the elements of $\{M_{AB}\}$ are placed into (1, 1) correspondence with those of $\{m\}$; it will then be shewn that an M_{AB} can always be found which is not included in the correspondence. Each M^0_{AB} in $\{M_{AB}\}$ now corresponds to a definite m_0 in $\{m\}$; form a new aggregate M'_{AB} in the following manner:—For each element M^0_{AB} in $\{M_{AB}\}$, in which A takes the place of m_0 in $\{m\}$, write B; and for each element M^0_{AB} in $\{M_{AB}\}$, in which B takes the place of m_0 in $\{m\}$, write A; in this manner we form an aggregate M'_{AB} in which each element is either A or B, which is similar to $\{m\}$, and which is not identical with any M_{AB} that occurs in the correspondence

* See *Jahresbericht d. Deutsch. Math. Vereinigung*, 1897.

between $\{M_{AB}\}$ and $\{m\}$. It has thus been shewn that the cardinal number of the aggregate of all the M_{AB} is greater than that of M. If M is the aggregate $a_1, a_2, a_3, \ldots a_n, \ldots$ which is similar to the aggregate of integral numbers, and if for A and B we write 0 and 1, then the aggregate $\{M_{01}\}$ may be interpreted as the aggregate of all the rational and irrational binary fractions; and this aggregate is thus shewn to be unenumerable. Instead of replacing the elements of $\{m\}$ by two letters A, B, we might have taken any finite number of letters without altering the principle of the proof. In § 56, the ten digits 0, 1, ... 9, were taken instead of A and B. It will be observed that, even if $\{m\}$ is normally ordered (see § 130), the new aggregate $\{M\}$ is not given as a normally ordered aggregate; and in default of proof it cannot be assumed that it is capable of being arranged in normal order.

To replace all the elements of an aggregate either by A or by B, is equivalent to taking a part * of the given aggregate. The theorem has thus been established that, *the cardinal number of the aggregate, each element of which is a part of a given aggregate, is greater than the cardinal number of the given aggregate, all possible parts being contained in the new aggregate.*

DIVISION OF CARDINAL NUMBERS BY FINITE NUMBERS.

121. If two aggregates have the same cardinal number, and if each of the two aggregates be divided into the same finite number n of parts, such that the n parts of the first aggregate all have the same cardinal number, and also the n parts of the second all have the same cardinal number, then it can be proved that the cardinal number of one of the parts of the first aggregate is the same as that of one of the parts of the second aggregate. Symbolically, the theorem may be stated in the form:—if α, β are cardinal numbers such that $n\alpha = n\beta$, then $\alpha = \beta$.

This theorem has been proved by Bernstein†. It will be sufficient to give the detailed proof in the case $n = 2$, as the proof in the general case is obtained by generalization of that employed in the particular case.

Since an aggregate is equivalent to itself, any special mode of exhibiting such equivalence, by which each element is made to correspond to a definite other element, is called a transformation of the system into itself. As regards all such possible transformations the following propositions may be seen to hold:—

(1) The transformations of an aggregate M into itself form a *group* ϕ_M.

(2) Let $1, \chi_1, \chi_2, \chi_3, \ldots$ denote a sequence of transformations of M into itself, 1 denoting the identical transformation, and let this sequence form a group which is necessarily a sub-group of ϕ_M; then the condition that the

* See Borel, *Leçons sur la théorie des fonctions*, p. 108.

† Inaugural Dissertation, "Untersuchungen aus der Mengenlehre," Halle, 1901. This is reproduced in *Math. Annalen*, vol. LXI.

sequence forms a group is that, corresponding to any two integers m, n, there is a third r, such that $\chi_m\chi_n = \chi_r$. Further, let us suppose that to every χ_n there corresponds a definite χ'_n, such that $\chi_n\chi'_n = 1$. If m be an element of M such that $m \neq \chi_n(m)$, for $n = 1, 2, 3, \ldots$, then $\chi_n(m) \neq \chi_{n'}(m)$, where n and n' are any unequal integers.

(3) If m and m' are any two distinct elements of M, and if $m \neq \chi_n(m')$, for $n = 1, 2, 3, \ldots$, then $\chi_n(m) \neq \chi_{n'}(m')$. For if $\chi_n(m) = \chi_{n'}(m')$, we should deduce that $m = \chi'\chi_n(m) = \chi'\chi_{n'}(m') = \chi_{n''}(m')$, which is contrary to the hypothesis made.

(4) If T_1, T_2, ... are parts of M, such that each one T has no element in common with another T, we may say that the T's form a system of separate parts of M.

If $T = \{t\}$ is a part of M, and if $t \neq \phi_n(t')$, for $n = 1, 2, 3, \ldots$, then the equivalent aggregates $T, \chi_1(T), \chi_2(T), \ldots$ form a system of separate parts of M.

(5) If T is a part of M which satisfies the condition stated in (4), then $\overline{\overline{M}} = \overline{\overline{M}} + \overline{\overline{T}}$.

For $T, \chi_1(T), \chi_2(T), \ldots$ are all parts of M having the cardinal number $\overline{\overline{T}}$; and if R is the part of M which remains when all these separate parts are removed, we have

$$\overline{\overline{M}} = \overline{\overline{R}} + \aleph_0 . \overline{\overline{T}};$$

hence

$$\overline{\overline{M}} + \overline{\overline{T}} = \overline{\overline{R}} + (\aleph_0 + 1)\overline{\overline{T}}$$
$$= \overline{\overline{R}} + \aleph_0 . \overline{\overline{T}} = \overline{\overline{M}}.$$

To proceed to the proof of the theorem:—Let

(a) $$\overline{\overline{M}} = \overline{\overline{x}}_1 + \overline{\overline{x}}_2 = \overline{\overline{x}}_3 + \overline{\overline{x}}_4,$$

(b) $$\overline{\overline{x}}_1 = \overline{\overline{x}}_2,$$

(c) $$\overline{\overline{x}}_3 = \overline{\overline{x}}_4;$$

then it is required to shew that $\overline{\overline{x}}_1 = \overline{\overline{x}}_3$, which involves $\overline{\overline{x}}_2 = \overline{\overline{x}}_4$.

The three equations (a), (b), (c) may be regarded as denoting that there are three reversible transformations of the aggregate M into itself, which may be denoted by ϕ_a, ϕ_b, ϕ_c respectively; the reversibility of these transformations is expressed by $\phi_a^2 = \phi_b^2 = \phi_c^2 = 1$.

The transformation ϕ_a involves $x_1 = (x_{13}, x_{14})$, where x_{13} are those elements of x_1 which are transformed into elements of x_3, and x_{14} those which are transformed into elements of x_4; on the whole we have

(6) $$\begin{cases} x_1 = (x_{13},\ x_{14}), \\ x_2 = (x_{23},\ x_{24}), \\ x_3 = (x_{31},\ x_{32}), \\ x_4 = (x_{41},\ x_{42}), \end{cases} \quad \text{where } \overline{\overline{x}}_{ik} = \overline{\overline{x}}_{ki}.$$

If T_1 is any part of x_1, and T_2 an equivalent part of x_2, we may denote by x_1^*, x_2^* the aggregates obtained by interchanging those elements of x_1 which belong to T_1 with those of x_2 which belong to T_2; we have then a similar set of equations to (6) for the new starred aggregates, and $\overline{\overline{x}}_1^* = \overline{\overline{x}}_1$, $\overline{\overline{x}}_2^* = \overline{\overline{x}}_2$, $\overline{\overline{x}}_3^* = \overline{\overline{x}}_3$, $\overline{\overline{x}}_4^* = \overline{\overline{x}}_4$. If then the theorem be proved for the starred aggregates, it holds for the original ones.

We have to shew that, after suitable transformations, a system of division of the aggregates into parts, of the form in (6), can be found, such that $\overline{\overline{x}}_{13} + \overline{\overline{x}}_{14} = \overline{\overline{x}}_{14}$, and $\overline{\overline{x}}_{31} + \overline{\overline{x}}_{32} = \overline{\overline{x}}_{32}$. For from these equations we deduce $\overline{\overline{x}}_1 = \overline{\overline{x}}_2 = \overline{\overline{x}}_{14}$, $\overline{\overline{x}}_3 = \overline{\overline{x}}_4 = \overline{\overline{x}}_{23}$, and then the aggregates x_3, x_4 are such that each has a part which is equivalent to the other; and consequently, in accordance with the equivalence theorem, x_3, x_4 are equivalent to one another; or $\overline{\overline{x}}_3 = \overline{\overline{x}}_4$. It has in fact to be shewn that x_{13} can be so chosen, that it is negligible with respect to cardinal number, in comparison both with x_{14} and with x_{23}.

We form the systems of transformation

$$\phi_b = \chi_2,\ \phi_b\phi_c = \chi_4,\ \phi_b\phi_c\phi_b = \chi_6,\ \dots,$$
$$\phi_c = \chi_3,\ \phi_c\phi_b = \chi_5,\ \phi_c\phi_b\phi_c = \chi_7,\ \dots;$$

each transformation χ in this system has one inverse, given by the scheme $\chi_{4n}\chi_{4n+1} = 1$, $\chi_{4n+2}\chi_{4n+2} = 1$, $\chi_{4n+3}\chi_{4n+3} = 1$; thus the transformations χ form a group of reversible transformations of $M = (x_1, x_2)$ into itself.

An element e_{13} of x_{13} is either, (i) transformed into an element of x_{24} by a transformation χ with finite index, or else, (ii) e_{13} is not transformed into an element of x_{24} by any of the transformations χ. Suppose, then, that for every element e_{13} of x_{13} the second of these cases arises, then χ_{2r}, χ_{2r+1} transform the elements of x_{13} into aggregates which are respectively in x_{23} and x_{14}, and in them these aggregates form an enumerable system of separate parts of each. For, in the case contemplated, χ_2 transforms x_{13} into a part of x_{23}; by χ_4, the elements of x_{13} become elements of x_{23} or x_{24}, consequently in accordance with (ii), $\chi_4(x_{13})$ is a part of x_{23}. In this manner it is seen, that x_{13} is transformed, by every χ_{2n}, into a part of x_{23}, and by every χ_{2n+1} into a part of x_{14}. It then follows, by Lemma (5), that $\overline{\overline{x}}_{13} + \overline{\overline{x}}_{14} = \overline{\overline{x}}_{14}$, $\overline{\overline{x}}_{13} + \overline{\overline{x}}_{23} = \overline{\overline{x}}_{23}$, and the theorem is then completely established. The remainder of the proof consists in shewing that, by an exchange of elements of x_{13} with elements of x_{24}, it is possible to arrange so that the case just considered always arises.

Suppose x_{13}' are those elements of x_{13} which are transformed by χ_2 into elements of x_{24}; let x_{13}'' denote those elements different from x_{13}' which are transformed by χ_3 into elements of x_{24} which were not affected by χ_2, and so on; we have then the scheme

$$\begin{cases} x_{13}' & \chi_2(x_{13}') & \text{in } x_{24}, \\ x_{13}' \neq x_{13}'' & \chi_2(x_{13}') \neq \chi_3(x_{13}'') & \text{in } x_{24}, \\ x_{13}' \neq x_{13}'' \neq x_{13}''' & \chi_2(x_{13}') \neq \chi_3(x_{13}'') \neq \chi_4(x_{13}''') & \text{in } x_{24}, \\ x_{13}' \neq \dots \neq x_{13}^{(n)} & \chi_2(x_{13}') \neq \chi_3(x_{13}'') \dots \neq \chi_n(x_{13}^{(n)}) & \text{in } x_{24}. \end{cases}$$

We take now the equivalent sums

$$[x_{13}] = \sum_{n=1}^{\infty} x_{13}^{(n)}, \text{ and } [x_{24}] = \sum_{n=1}^{\infty} \chi_{n+1}(x_{13}^{(n)}),$$

and we carry out an exchange of $[x_{13}]$ with $[x_{24}]$; we then have

$$x_{13} = [x_{13}] + [(x_{13})], \quad x_{24} = [x_{24}] + [(x_{24})].$$

When the exchange has been made of the elements of $[x_{13}]$ with those of $[x_{24}]$, we denote the new aggregates by starring the original ones; we have then, in accordance with the formulae (6), expressions for x_1^*, x_2^*, x_3^*, x_4^*, and we can, as has been shewn above, attend to these, instead of to the original x_1, x_2, x_3, x_4. Now no element of x_{13}^* is transformed into an element of x_{24}^* by any of the transformations χ, it being understood that the transformations χ are not to affect the substituted elements; and thus by the reasoning which has been given above for the case in which no element of x_{13} is transformed into an element of x_{24}, the theorem is established. Bernstein has also proved that if $2\alpha = \alpha + \beta$, where α, β are cardinal numbers, then $\alpha \geqq \beta$.

THE ORDER-TYPE OF SIMPLY ORDERED AGGREGATES.

122. *An aggregate M is said to be a simply ordered aggregate when each element m has a definite rank relatively to the other elements of M, so that, of any two elements m, m′ whatever, it is known which has the higher and which has the lower rank.*

If m has a lower rank than m', the fact is denoted symbolically by $m < m'$; and if a higher rank, by $m > m'$.

If an aggregate is given at first unordered, it may be possible to order the aggregate in a variety of essentially distinct ways. If the aggregate is finite, the ordering of it may be accomplished by arbitrarily assigning to each element its rank relatively to the others. In case the aggregate is an infinite one, the ordering of it consists in the setting up of some general rule which suffices logically to assign the relative order of any two elements.

Besides simply ordered aggregates there exist also doubly or trebly ordered aggregates, or also aggregates with higher degrees of multiplicity of order. Each element of such an aggregate possesses two, three, or more distinct characteristics of an ordinal character. Simply ordered aggregates only will be here considered.

Two simply ordered aggregates M, N are said to be similar, when a (1, 1) *correspondence can be established, in accordance with some law, such that to any two definite elements m, m′ of M there correspond two definite elements n, n′* of *N, in such a manner that the relative order of m, m′ in M, is the same as that of the corresponding elements n, n′ in N.*

This relation of similarity may be represented symbolically by $M \simeq N$.

Every simply ordered aggregate is similar to itself.

Two simply ordered aggregates which are similar to a third are similar to one another.

All simply ordered aggregates which are similar to one another are said to have the same order-type.

An order-type is accordingly characteristic of a class of similar aggregates.

The order-type of a simply ordered aggregate M is defined by Cantor as the concept which is obtained by abstraction when the nature of the elements of M is disregarded, their order being alone retained. The order-type of M is then denoted by $\bar{M}$. This definition will be further discussed in § 155. That similar aggregates have the same order-type is regarded by Cantor as a deduction from this definition.

If in $\bar{M}$, we further disregard the order of the elements, we obtain $\bar{\bar{M}}$, the cardinal number of M.

The order-type of M is, from Cantor's point of view, regarded as a simply ordered aggregate similar to M, such that each element is the number 1. If any order-type be denoted by α, the corresponding cardinal number is denoted by $\bar{\alpha}$.

Corresponding to any given transfinite cardinal number, there is a multiplicity of order-types, which form a class of order-types; each such class of order-types is characterised by the common cardinal number of all the order-types of the class.

The order-types which belong to the class corresponding to a cardinal number α, form an aggregate which has a cardinal number α'. It will appear that α' is always greater than α.

If the order of every pair of elements in a simply ordered aggregate M be reversed, the aggregate in the new order is denoted by $^*M$.

If $\bar{M} \equiv \alpha$, then the order-type $^*\bar{M}$ is denoted by $^*\alpha$.

The order-type of the aggregate of all the finite integers in their natural order $(1, 2, 3, \ldots)$, is denoted by ω. This is therefore the order-type of every aggregate $(a_1, a_2, \ldots a_n \ldots)$ which is similar to $(1, 2, 3, \ldots)$.

The aggregate $(\ldots a_n \ldots a_3, a_2, a_1)$ has the order-type $^*\omega$.

THE ADDITION AND MULTIPLICATION OF ORDER-TYPES.

123. If M, N denote two simply ordered aggregates, and if the aggregate (M, N) be formed, in which all the elements of both M and N occur, and which is such that any two elements of M have the same relative order as in M, and that any two elements of N have the same relative order as in N, and further that each element of M has a lower rank than all the elements of N, then the new simply ordered aggregate (M, N) is said to be the *sum* of the two simply ordered aggregates M and N. It is clear that if $M \simeq M'$,

$N \simeq N'$, then $(M, N) \simeq (M', N')$, and thus that the order-type of (M, N) depends only on the order-types of M and N.

If $\bar{M} = \alpha$, $\bar{N} = \beta$, the sum $\alpha + \beta$ is defined to be the order-type of the sum (M, N) of the two simply ordered aggregates, as defined above.

This defines the operation of addition of order-types. It will be seen that the addition of order-types does not obey the commutative law. For if $\alpha = \bar{M}$, $\beta = \bar{N}$, then $\alpha + \beta = \overline{(M, N)}$: but $\beta + \alpha = \overline{(N, M)}$; and the two order-types $\overline{(M, N)}$, $\overline{(N, M)}$ are in general different from one another.

If n denotes a finite integer, $\omega + n$ is the order-type of the ordered aggregate $(e_1, e_2, e_3, \ldots f_1, f_2, \ldots f_n)$, whereas $n + \omega$ is the order-type of $(f_1, f_2, \ldots f_n, e_1, e_2, e_3, \ldots)$. It is clear that the first of these aggregates is not similar to $(g_1, g_2, g_3 \ldots)$, but if we let $f_1, f_2, \ldots f_n$ correspond to $g_1, g_2, g_3 \ldots g_n$, then e_1 to g_{n+1}, e_2 to $g_{n+2}, \ldots$ and in general e_m to g_{n+m}, it is seen that the second of the above order-types is similar to $(g_1, g_2, g_3 \ldots)$. It thus appears that $n + \omega = \omega$, but $\omega + n \neq \omega$.

124. In the simply ordered aggregate N, let us suppose that in the place of each element is substituted a simply ordered aggregate similar to M, whereby a new simply ordered aggregate is formed; this may be denoted by $M . N$. It is clear that if $M \simeq M'$, $N \simeq N'$, then $M . N \simeq M' . N'$, thus the order-type of $M . N$ depends only on the order-types of M and N.

If $\alpha = \bar{M}$, $\beta = \bar{N}$, the product $\alpha . \beta$ is defined to be $\overline{M . N}$, the order-type of $M . N$, as just defined.

It will be seen that the product $\alpha . \beta$ is in general different from $\beta . \alpha$, and thus that the multiplication of order-types does not obey the commutative law. For example $\omega . 2$, is the order-type of the aggregate formed by substituting in (a_1, a_2) for each of the two elements an aggregate of type ω; $\omega . 2$ is therefore the order-type of $(b_1, b_2, b_3, \ldots c_1, c_2, c_3 \ldots)$, in which there is no last element, and no element immediately preceding c_1. On the other hand, $2 . \omega$ is the order-type obtained by substituting for each element in $(a_1, a_2, a_3 \ldots)$, an aggregate consisting of two elements; and $2 . \omega$ is thus the order-type of the enumerable aggregate $(a_{11}, a_{12}, a_{21}, a_{22}, a_{31}, a_{32} \ldots)$, which is similar to $(b_1, b_2, b_3 \ldots)$, as may be seen by making a_{n1} correspond to b_{2n-1} and a_{n2} to b_{2n}. It has thus been shewn that $2 . \omega = \omega$, but $\omega . 2 \neq \omega$.

THE STRUCTURE OF SIMPLY ORDERED AGGREGATES.

125. An examination of the structure of a simply ordered aggregate M can, in general, only be attempted by considering the nature of those aggregates which are its parts, and in each of which parts the order of the elements is the same as that of the same elements in the whole aggregate. The simplest transfinite part of an ordered aggregate is that which has one of the types ω, $^*\omega$. Such parts we speak of as *ascending sequences*, and *descending sequences*, respectively, contained in M.

Two ascending sequences $\{a_n\}$, $\{a_n'\}$, contained in M, are said to be *related to one another*, provided that, corresponding to any element a_n of the first, there are elements $a'_{n'}$ of the second, such that $a_n < a'_{n'}$; and provided also that, corresponding to any element a_n' of the second, there are elements $a_{n'}$ of the first sequence, such that $a_n' < a_{n'}$.

Two descending sequences $\{b_n\}$, $\{b_n'\}$ contained in M, are said to be related to one another, provided that, corresponding to any element b_n of the first sequence, there are elements $b'_{n'}$ of the second, such that $b_n > b'_{n'}$; and provided also that, corresponding to any element b_n' of the second, there are elements $b_{n'}$ of the first sequence, such that $b_n' > b_{n'}$.

An ascending sequence $\{a_n\}$, and a descending sequence $\{b_n\}$, contained in M, are said to be related to one another, if $a_n < b_{n'}$, for every n and n'; and further, provided there exists in M no element, or only one element m, which is such that $a_n < m < b_n$, for every n.

Two sequences contained in an ordered aggregate, which are both related to a third sequence, are related to one another.

Two sequences in an ordered aggregate, which are both ascending, or both descending, and of which one is a part of the other, are related to one another.

126. Suppose that in an ordered aggregate M, there is an element m_0 which satisfies the following conditions, with respect to an ascending sequence contained in M:

(1) for every n, $a_n < m_0$;

(2) for every element m of M which is $< m_0$, there exists a number n such that $a_n, a_{n+1}, a_{n+2}, \ldots$ are all $> m$; then the element m_0 is said to be the *limiting element*, or *limit of* $\{a_n\}$ *in* M; and m_0 is said to be a *principal element* of M.

Similarly, if we suppose that in M, there is an element m_0, which satisfies with reference to a descending sequence $\{a_n\}$ contained in M, the following conditions:

(1) for every n, $a_n > m_0$;

(2) for every element m of M which is $> m_0$, there exists a number n such that $a_n, a_{n+1}, a_{n+2}, \ldots$ are all $< m$; then the element m_0 is said to be a *limiting element*, or *limit of* $\{a_n\}$ *in* M; and m_0 is said to be a *principal element* of M.

A sequence contained in M can never have more than one limiting element in M.

If a sequence in M has a limiting element m_0 in M, then m_0 is the limiting element of every sequence in M which is related to the first one.

Two sequences which have the same limiting element in M, must be related to one another.

It is clear that, if M, M' are similar ordered aggregates, an ascending or a descending sequence in M corresponds to a sequence of the same kind in M'. To every principal element in M, there corresponds a principal element in M'.

An ordered aggregate which is such that every element is a principal element is said to be dense-in-itself.

If, in an ordered aggregate, every sequence which is contained therein has a limiting element in the aggregate, then the ordered aggregate is said to be a closed aggregate.

An ordered aggregate which is dense-in-itself, and also closed, is said to be perfect.

An ordered aggregate which is such that between any two whatever of its elements, there are other elements of the aggregate, is said to be everywhere-dense.

The properties of an ordered aggregate thus defined, are also properties of any similar aggregate; hence the terms may be applied to the order-types which are symbolised by replacing the elements of the ordered aggregates by 1; there can exist therefore an order-type which is dense-in-itself, or closed, or perfect, or everywhere-dense.

The terms which have been here employed for the purpose of describing certain peculiarities which may exist in an ordered aggregate, or in the corresponding order-type, are identical with those which we have employed in analogous senses in Chapter II, in the case of sets of points or numbers. There is however a distinction which must be noticed between the use of the terms in the two cases. To illustrate this distinction, let $(P_1, P_2, P_3 \ldots P_n \ldots)$ be a sequence of points on a straight line, which sequence has a limiting point P_ω on the right of the points P_n; then if Q be any point of the straight line on the right of P_ω, the two ordered aggregates $(P_1, P_2, P_3, \ldots P_n, \ldots P_\omega)$, and $(P_1, P_2, P_3, \ldots P_n, \ldots Q)$, are similar, and have the same order-type $\omega + 1$. In the first of these aggregates, P_ω is the limiting element of the sequence $(P_1, P_2, \ldots P_n \ldots)$; and, in the second aggregate, Q is the limiting element of the same sequence; and therefore both the ordered aggregates are closed, in the sense explained above. The first of these aggregates forms a closed set of points, in the sense of the term defined in Chapter II; but the second does not, since Q is not a limiting point of the set of points $\{P_n\}$. The distinction rests upon the different use of the terms limiting element and limiting point, in the two cases of an ordered aggregate of elements in general, and that of a set of points in the continuum. The question whether an element is a limiting element of an aggregate to which it belongs, or not, in the sense defined above, is answered by examining the structure of the ordered aggregate itself. In the case of a set of points in the continuum, a particular point may be a limiting element of the aggregate of points considered merely as an aggregate of elements with a particular order-type; but the question as to whether the same point is a limiting point of the set of points, considered as chosen out of the continuum, can only be answered after an examination of the ordinal relation of the point to other points of the continuum which do not belong to the set; in fact, the set must be regarded, for this purpose, as an aggregate which is only a part of another aggregate, the continuum. It is now clear that a set of points considered solely as an ordered aggregate of elements, without reference to the fact that it is essentially a part of the continuum, may be closed, or perfect; and yet that the same set of points need be neither closed nor perfect, in the sense of the terms employed in the theory of sets of points, which has been dealt with in Chapter II.

THE ORDER-TYPES η, θ, π.

127. Certain order-types which are of special importance will be now examined.

The first of these is the order-type η, of the set R of rational numbers between 0 and 1 (both exclusive), in their order as defined in Chapter I.

It will be shewn that the order-type η, is exhaustively characterised by the following properties:—

(1) $\bar{\eta} = \aleph_0 = a$.

(2) In η, there is no lowest and no highest element.

(3) η is everywhere-dense.

In fact, every simply ordered aggregate M, which has these three characteristics, is similar to the aggregate R.

To prove this, we first observe that, on account of the condition (1), the order of the elements in both M and R can be so altered that each of them is reduced to the order-type ω. Let this be done; and denote by M_0, R_0 the new ordered aggregates

$$M_0 = (m_1, m_2, m_3 \ldots),$$
$$R_0 = (r_1, r_2, r_3 \ldots).$$

We have to shew that $M \simeq R$; and to do this we have to shew how to establish the requisite correspondence between the elements m of M, and r of R. Let m_1 be made to correspond to r_1; then there are an indefinitely great number of elements of M, which have the same relation, as regards order, to m_1, as r_2 has, in R, relatively to r_1; of all these elements choose that one m_{ϵ_2}, which has the smallest index as it appears in M_0; and let m_{ϵ_2} be made to correspond to r_2. Of all the elements of M, which are related to m_1 and m_{ϵ_2}, in the same manner, as regards order in M, as r_3 is related to r_1 and r_2, as regards order in R, choose that one m_{ϵ_3}, which has the smallest index as it appears in M_0; and make m_{ϵ_3} correspond to r_3.

Proceeding in this manner, we make the elements $r_1, r_2, r_3 \ldots r_n$ of R, correspond to the elements $m_1, m_{\epsilon_2}, m_{\epsilon_3} \ldots m_{\epsilon_n}$ of M; and so far as these elements are concerned the relations of rank are preserved in the correspondence: we proceed then to choose, in the same manner as before, the element $m_{\epsilon_{n+1}}$, which is to be made to correspond to r_{n+1}; and thus we obtain, for every r_n, the corresponding m_{ϵ_n}. It must however be shewn that this process exhausts all the elements m of M, that is to say, that in the sequence $1, \epsilon_2, \epsilon_3, \ldots \epsilon_n, \ldots$ every integral number p occurs in some definite place. This can be proved by the method of induction. Let us assume that the elements $m_1, m_2, m_3, \ldots m_n$ all occur in the correspondence that has been set up between the whole of R and at least a part of M, then we shall prove that m_{n+1} also occurs. Upon this assumption, let λ be so great that among the elements $m_1, m_{\epsilon_1}, m_{\epsilon_2}, \ldots m_{\epsilon_\lambda}$, all the elements $m_1, m_2, m_3, \ldots m_n$ occur. Then if m_{n+1} is not also among those elements, choose out of $r_{\lambda+1}, r_{\lambda+2}, r_{\lambda+3}, \ldots$ that element $r_{\lambda+s}$ with the smallest index which has the same relation to $r_1, r_2, \ldots r_\lambda$, as regards order in R, that m_{n+1} has relatively to $m_1, m_{\epsilon_1}, m_{\epsilon_2}, \ldots m_{\epsilon_\lambda}$, as regards order in M. Then the element m_{n+1} has the same relation to $m_1, m_{\epsilon_1}, m_{\epsilon_2}, \ldots m_{\epsilon_{\lambda+s-1}}$, as regards order in M, as $r_{\lambda+s}$ has to $r_1, r_2, \ldots r_{\lambda+s-1}$, as regards order in R. It thus appears that m_{n+1} is the element with the smallest index as it appears in M_0, which has, in M, the same relation as regards order to $m_1, m_{\epsilon_2}, \ldots m_{\epsilon_{\lambda+s-1}}$, that $r_{\lambda+s}$ has relatively to $r_1, r_2, \ldots r_{\lambda+s-1}$ in R; hence $m_{\epsilon_{\lambda+s}} \equiv m_{n+1}$; that is, the element m_{n+1} occurs in the correspondence which has been established between M and R. It has now been shewn that M and R are similarly ordered aggregates.

Examples of the order-type η are the following:—

(1) The aggregate of all negative and positive rational numbers including zero, in their natural order.

(2) The aggregate of all rational numbers which are greater than a, and less than b, where a, b are two real numbers such that $a < b$.

(3) The aggregate of all real algebraical numbers in their natural order in the continuum, or of all such of these numbers as lie between two real numbers a, b.

(4) The aggregate of a set of non-abutting linear intervals which are such that their end-points and the limiting points of these end-points form a non-dense perfect set of points in a linear interval.

The rational numbers of the interval (0, 1), including 0 and 1, form an aggregate of the order-type $1 + \eta + 1$.

128. We now proceed to the consideration of the order-type θ, of points forming a linear continuum.

It will be shewn that any simply ordered aggregate M is similar to the aggregate X of all real numbers of the continuum $(0, 1)$, in their natural order, provided (1) M is perfect, and (2) in M, an aggregate S, with the cardinal number $\aleph_0$, is contained, which is so related to M, that, between any two elements m_0, m_1 of M, there are elements of S.

If S has a lowest and a highest element, these can be removed without affecting its relation to M; and thus we may suppose S to be of the type η, of the aggregate R of rational numbers which lie between 0 and 1, both exclusive, in their natural order.

Since $S \simeq R$, we may suppose the elements of S to be made to correspond in order to the elements of R; and it will be shewn that this correspondence enables us to establish a correspondence between the elements of M and of X.

We suppose that each element of M, which belongs to S, corresponds to that element of X which belongs to R, just as in the correspondence of S with R already established. Any element m of M, which does not belong to S, is the limiting element of a sequence $\{m_n\}$ of elements of S. To this sequence $\{m_n\}$, there corresponds a sequence $\{r_n\}$ in X, all the elements of which belong to R; and this sequence $\{r_n\}$ has a limiting element x in X not belonging to R; we take therefore m in M to correspond to x in X. If we take a different sequence $\{x_n'\}$, which has the same limiting element m as before, in M, then there corresponds to it a sequence $\{r_n'\}$ in R, which has the same limiting element x as before, in X. It will now be shewn that, in the correspondence so established between the elements of M and of X, the relative order of two elements of M is the same as that of the corresponding elements of X. This clearly holds of any two elements of M which are also elements of S. Consider next two elements m and s, of M, the first of which does not, and the second of which does, belong to S; and let x_1, r be the corresponding elements of X. If $r < x_1$, there exists an ascending sequence in R, of which x_1 is the limiting element, such that all its elements are $> r$; then to this sequence there corresponds an ascending sequence in S, all the elements of which are $> s$, and of which m is the limiting element; hence $s < m$. If $r > x_1$, it can, in a similar manner, be shewn that $s > m$. The proof that, corresponding to any two elements m_1, m_2 of M which do not belong to S, the elements x_1, x_2 of X are such that $m_1 \gtrless m_2$, according as $x_1 \gtrless x_2$, is of a precisely similar character to that just given. It has thus been shewn that M and X are similar aggregates, and that the type θ is characterised by the conditions (1) and (2).

The above characterisation* of the type θ contains Cantor's ordinal theory of the constitution of the linear continuum.

* See Russell, *Principles of Mathematics*, vol. I, p. 303, also Veblen, *Trans. Amer. Math. Soc.*, vol. VI, and Huntington, *Annals of Math.*, Ser. 2, vols. VI and VII.

A non-dense perfect set of points in a linear interval has not the order-type θ, but the set of complementary intervals together with the limiting points of their end-points does form an aggregate of order-type θ, when the elements consisting partly of points and partly of intervals are taken in the order in which they occur in the continuum.

129. The order-type $^*\omega + \omega$ may be denoted by π, and is the order-type of the negative and positive integers in their natural order. This order-type has properties distinct from that of ω. For example, $n + \omega$ has been shewn to be identical with ω, where n is a finite integer, but $n + \pi$ is not identical with π. From either of the equations $n + \pi = m + \pi$, or $\pi + n = \pi + m$, there follows $m = n$, or more generally†:—

If n, n' are finite integers, ζ and ζ' other order-types, from the equation $n + \pi + \zeta = n' + \pi + \zeta'$, there follows $n = n'$, $\zeta = \zeta'$.

To prove this theorem, we observe that, if the two aggregates be placed into similar correspondence, the lowest elements correspond to one another, then the second, and so on; hence $n = n'$ is proved at once: and we now have $\pi + \zeta = \pi + \zeta'$.

When two simply ordered aggregates $M_\pi + Z$, $N_\pi + Z'$ of order-types $\pi + \zeta$, $\pi + \zeta'$ are placed in correspondence in order, either M_π corresponds to N_π, or M_π corresponds to a part of N_π, or else N_π corresponds to a part of M_π. In the last two cases the order-type π must be split up into $\pi = \pi_1 + \pi_2$, where $\pi = \pi_1$, and π_2 is some other order-type; but from the definition $\pi = {}^*\omega + \omega$, it is clear that every mode of dividing π into two parts without altering the relative order of the elements, leaves it in the form $^*\omega + \omega$; hence it is impossible that $\pi = \pi_1 + \pi_2$, and $\pi = \pi_1$, and therefore M_π corresponds to N_π. Hence also Z corresponds to Z', or $\zeta = \zeta'$.

NORMALLY ORDERED AGGREGATES.

130. The order-type of a simply ordered aggregate is, as we have already seen, such that the structure of the aggregate as revealed by an examination of the sequences contained in it, may be of the most varied character; the various sequences may be ascending or descending ones, and may or may not have a limiting element within the aggregate.

Of all the possible order-types, those are of especial importance which have been defined by Cantor as the order-types of normally ordered aggregates (wohlgeordnete Mengen).

A normally ordered aggregate M is one which satisfies the following conditions:—

(1) *M has an element m, of lower rank than all the other elements.*

† Bernstein, *loc. cit.*, p. 9.

(2) *If M_1 is any part of M, and if M contains one or more elements which are of higher rank than all the elements of M_1, then there exists one element m' of M, which immediately follows the part-aggregate M_1, so that there are no elements of M which are intermediate in rank between m' and all the elements of M_1.*

The special case of (2) which arises when M_1 consists of one element, shews that a normally ordered aggregate is such that each element has one which immediately follows it, unless the element is the highest element of M. It is however not necessarily the case that M has a highest element.

If $e_1, e_2, e_3, \ldots e_n, \ldots$ is an ascending sequence of elements contained in M, and such that elements exist in M, which are of higher rank than every e_n, then there exists an element e' of M which is higher than all the e_n, and such that every element e'' of M which is lower than e' is lower than $e_n, e_{n+1}, e_{n+2}, \ldots$ for some definite value of n.

Every part of a normally ordered aggregate has a lowest element.

Let M_1 be a part of M; if M_1 contains m_1 the lowest element of M, then m_1 is the lowest element of M_1. If M_1 does not contain m_1, consider that part of M which contains all those elements every one of which is of lower rank than all the elements of M_1; this part of M must have an element which immediately follows it; and this element belongs to M_1, and is its lowest element.

If a simply ordered aggregate M itself, and also every part of M, has a lowest element, M is normally ordered.

The condition (1) is fulfilled. Let M_1 be a part of M such that M contains elements which are higher than all those of M_1; let these form the aggregate M_2, and let m be the lowest element of M_2. Then m is the element which immediately follows M_1; and thus the condition (2) is satisfied.

This property of a normally ordered aggregate, that every part of it has a lowest element, might be adopted as the definition of a normally ordered aggregate.

A somewhat simpler property which might be employed to define a normally ordered aggregate is the following:—

An aggregate M is normally ordered† *if, and only if, it contains no part of which the order-type is* $^*\omega$.

If M is not normally ordered at least one part of it must have no lowest element, and this part contains a sequence whose order-type is $^*\omega$. An aggregate which has a lowest element, and is also such that each element has one that immediately succeeds it, is not necessarily normally ordered,

† See Jourdain, *Phil. Mag.*, Ser. 6, vol. VII, p. 65.

even if each element has one immediately preceding it. This can be seen by considering an aggregate with the order-type $\omega + {}^*\omega$.

131. The following properties of normally ordered aggregates can be proved in a very simple manner:—

Every part-aggregate of a normally ordered aggregate is itself normally ordered.

Every ordered aggregate which is similar to a normally ordered aggregate is itself normally ordered.

If in a normally ordered aggregate M there be substituted for the elements normally ordered aggregates, in such a manner that if M_m, $M_{m'}$ are the aggregates substituted for any two elements m, m', then $M_m \lesseqgtr M_{m'}$, according as $m \lesseqgtr m'$, the resulting new aggregate is normally ordered.

132. *The part of a normally ordered aggregate M which consists of all those elements which are of lower rank than an element m, of M, is called the segment of M determined by the element m.*

The aggregate which remains when the segment of M, determined by the element m, is removed from M, is called the *remainder* of M determined by the element m. The element m is the lowest element of the remainder.

If S is the segment of M formed by m, and R is the remainder, then $M = (S, R)$.

Of two segments S, S' determined by the elements m, m' of which $m < m'$, we say that S is the smaller and S' the larger segment, or $S < S'$.

It can easily be seen that, if M, M_1 are two similar normally ordered aggregates, a segment of M corresponds to a similar segment of M_1, the element by which the segment of M is determined corresponding to the element of M_1 by which the segment of M_1 is determined.

A normally ordered aggregate is not similar to any of its segments.

Assume that, if possible, $S \simeq M$, and suppose the elements of S, M are put into correspondence. To the segment S of M, there must correspond a segment S_1 of S, so that $S_1 \simeq M \simeq S$, where $S_1 < S$. Since $S_1 \simeq M$, we find in a similar manner a segment $S_2 < S_1$, which is similar to M, and so on; and in this way we obtain an unending sequence $S > S_1 > S_2 \ldots > S_n \ldots$ of segments of M which are all similar to M. Let $m, m_1, m_2 \ldots m_n \ldots$ be the elements which determine the segments $S, S_1, S_2, \ldots S_n \ldots$; then $m > m_1 > m_2 \ldots > m_n \ldots$.

The aggregate $(\ldots\ldots m_n, \ldots m_2, m_1, m)$ would be a part of M which has no lowest element, and is of type ${}^*\omega$, which is impossible if M is normally ordered.

If M is an infinite normally ordered aggregate, it always has parts which are similar to M, although such a part cannot be a segment.

A normally ordered aggregate cannot be similar to any part of one of its segments.

Let us assume that, if possible, S' a part of a segment S, of M, is similar to M. Since $S' \simeq M$, we can place the elements of S', M in correspondence, then to the segment S of M there will correspond a segment S_1 of S', where $S_1 \simeq S$; let then S_1 be determined by the element e_1 of S'. Since e_1 is also an element of M, it determines a segment M_1 of M, of which S_1 is a part, and which has a part similar to M. Proceeding in the same manner, we determine a segment M_2 of M_1 which has a part that is similar to M; and in this way we obtain an unending sequence of segments of M, all similar to M, so that $M > M_1 > M_2 \ldots > M_n \ldots$. The elements which determine these sequences form a part of M which is of type $^*\omega$, and this is contrary to the hypothesis that M is normally ordered.

Two different segments of a normally ordered aggregate cannot be similar.

For one of these segments is a segment of the other.

There is only one mode of putting the elements of two similar normally ordered aggregates into correspondence, so that the relative orders of the elements are unaltered in the correspondence.

For if in two modes of placing the aggregates in correspondence two elements f, f' of one aggregate M, correspond to one element e of the other M', the segments of M determined by f, f' are each similar to the segment of M' determined by e; but it has been shewn to be impossible that M can have two different segments which are similar to one another.

A segment of one of two normally ordered aggregates has at most one segment of the other aggregate which is similar to it.

If S, S' are similar segments of two normally ordered aggregates M, M', then to every smaller segment $S_1 < S$, of M, there corresponds a similar segment $S_1' < S'$, of M'.

If S_1, S_2 are two segments of the normally ordered aggregate M, and S_1', S_2' are two similar segments of a normally ordered aggregate M', then if $S_1' < S_1$, it follows that $S_2' < S_2$.

If a segment S of M is not similar to any segment of another normally ordered aggregate M', then no segment $S' > S$ of M is similar to any segment of M' nor to M' itself; and the same holds of M itself.

If M, M', two normally ordered aggregates are so related that, to any segment of either, there corresponds a similar segment of the other, then $M \simeq M'$.

Any element e of M determines a segment of M which corresponds to a similar segment of M'. Let this latter be determined by an element e' of M'; we then take e to correspond to e'. To every element of M we therefore find a corresponding element of M', and it is seen by applying the foregoing theorems that the relative order of the elements is preserved.

133. *If two normally ordered aggregates M, M' are so related that,* (1) *to every segment S of M, there corresponds a similar segment S' of M', and* (2) *at least one segment of M' exists to which there is no corresponding similar segment of M; then there exists a definite segment S_1' of M' such that $S_1' \simeq M$.*

Consider all those segments of M', which do not correspond to similar segments of M. Among these, there must be one S_1' which is the least of all; this follows from the fact that the elements which determine these segments of M' form an aggregate which has a lowest element, and this lowest element determines the segment S_1'. Every segment of M' which is greater than S_1', is such that there exists no corresponding similar segment of M; but every segment of M' which is less than S_1' has a corresponding similar segment of M. Since to every segment of M there corresponds a similar segment of S_1', and to every segment of S_1' there corresponds a similar segment of M, it follows that $M \simeq S_1'$.

If the normally ordered aggregate M' has at least one segment to which there corresponds no similar segment of M, then to every segment of M there corresponds a similar segment of M'.

Let S_1' be the smallest segment of M', to which there corresponds no similar segment of M. If there existed segments of M to which no corresponding similar segments of M' exist, let S_1 be the smallest of all such segments of M. To every segment of S_1 there corresponds a similar segment of S_1', and conversely; hence $S_1 \simeq S_1'$, which is contrary to the hypothesis that there exists no segment of M which is similar to S_1'.

If M, M' are any two normally ordered aggregates, then either (1) M and M' are similar, or, (2) there exists a segment S' of M' which is similar to M, or, (3) there exists a segment S of M, which is similar to M', and these possibilities are mutually exclusive.

The following four possibilities may be contemplated, as regards the relation of M to M':—

(1) To every segment of either M or M' there corresponds a similar segment of the other aggregate.

(2) To every segment of M there exists a corresponding similar segment of M'; but there is at least one segment of M' to which no similar segment of M corresponds.

(3) To every segment of M' there corresponds a similar segment of M; but there is at least one segment of M to which no similar segment of M' corresponds.

(4) There is at least one segment of M to which no similar segment of M' corresponds, and also at least one segment of M' to which no similar segment of M corresponds.

It has been shewn that (4) is impossible. In the case (1), it has been proved that $M \simeq M'$. In the case (2), it has been shewn that a definite segment S_1' of M' exists, such that $S_1' \simeq M$; and in the case (3), that there is a definite segment S_1 of M such that $S_1 \simeq M'$.

It is impossible that at the same time $M \simeq M'$, and also $M \simeq S_1'$: for, in that case, $M' \simeq S_1'$; and it has been shewn to be impossible that M' is similar to one of its own segments.

It is also impossible that $M \simeq S_1'$, and also $M' \simeq S_1$; for there must then exist a segment of S_1', which is similar to S_1, and therefore to M'; but this is contrary to the theorem that a normally ordered aggregate cannot be similar to one of its segments.

If any part of M is such that that part is not similar to any segment of M, then that part is similar to M itself.

Any part M_1 of M is normally ordered; if then M_1 be similar neither to M nor to any segment of M, there must exist a segment of M_1' of M_1 which is similar to M; and M_1' is a part of that segment of M which is determined by the same element that determines the segment M_1' of M_1. Therefore M_1 would be similar to a part of one of its segments, which has been shewn to be impossible.

THE THEORY OF ORDINAL NUMBERS.

134. *The order-type $\bar{M}$ of a normally ordered aggregate M, is said to be the ordinal number which belongs to M; all similar normally ordered aggregates have consequently the same ordinal number.*

If M, M′ are two normally ordered aggregates such that M has a segment which is similar to M′, whilst M′ has no segment which is similar to M, then the ordinal number $\alpha \equiv \bar{M}$, is said to be greater than the ordinal number $\beta \equiv \bar{M}'$; and this relation is denoted by $\alpha > \beta$. If M has no segment similar to M′, but M′ has a segment similar to M, the ordinal number α is said to be less than β, and the relation is denoted by $\alpha < \beta$.

It follows from these definitions in conjunction with the theorem of § 133 that if α, β are any two ordinal numbers whatever, they satisfy one, and one only, of the relations $\alpha = \beta$, $\alpha > \beta$, $\alpha < \beta$; and that if $\alpha > \beta$, then $\beta < \alpha$.

Further it is seen that if $\alpha < \beta$ and $\beta < \gamma$, then $\alpha < \gamma$; hence the aggregate of all ordinal numbers is a simply ordered aggregate, when arranged in such a manner that any one α, which has been defined as less than another one β, precedes it.

The sum $\alpha+\beta$ of two ordinal numbers is, in accordance with the general definition of the sum of two order-types, the order-type of the normally ordered aggregate (M, N), where M, N are two normally ordered aggregates such that $\alpha=\bar{M}$, $\beta=\bar{N}$.

Since M, N each contains no part of type $^*\omega$, the same is true of (M, N). Hence the aggregate (M, N) is normally ordered; and thus $\alpha+\beta$ is an ordinal number.

Since M is a segment of (M, N), we see that $\alpha < \alpha+\beta$.

N is a remainder of (M, N) determined by the lowest element of N, hence N may be similar to (M, N); or, if not, it is similar to a segment of (M, N): thus either $\beta=\alpha+\beta$, or $\beta<\alpha+\beta$.

The addition of ordinal numbers obeys the associative law, but not in general the commutative law; thus $(\alpha+\beta)+\gamma=\alpha+(\beta+\gamma)$, but $\alpha+\beta$ is in general $\neq\beta+\alpha$.

135. The product $\alpha.\beta$ of two ordinal numbers is, in accordance with the definition of § 124, the order-type of the aggregate obtained by substituting for each element of an aggregate of order-type β, an aggregate of order-type α. In accordance with the theorem of § 131, the aggregate thus obtained is normally ordered, and of type dependent only on α and β.

In general $\alpha.\beta$ is not equal to $\beta.\alpha$.

It is easily seen that $\alpha.\beta>\alpha$, provided $\beta>1$; and that if $\alpha\beta=\alpha\gamma$, then $\beta=\gamma$.

If α, β are two ordinal numbers such that $\alpha<\beta$, there exists an ordinal number γ such that $\alpha+\gamma=\beta$; and this number γ is defined to be $\beta-\alpha$.

For if $\bar{M}=\beta$, there is a segment of M which may be denoted by M_1, such that $\bar{M}_1=\alpha$; let then $M=(M_1, S)$, therefore $\bar{M}=\bar{M}_1+\bar{S}$, and $\beta-\alpha=\bar{S}$.

136. Let $\beta_1, \beta_2, \dots \beta_n, \dots$ denote a simple sequence of ordinal numbers, and suppose $M_1, M_2, \dots M_n, \dots$ are aggregates of which the order-types are respectively the numbers of the sequence. The aggregate $(M_1, M_2, \dots M_n, \dots)$, which is obtained by replacing each element of the normally ordered aggregate $(1, 1, 1, \dots)$ of type ω, by a normally ordered aggregate, is, in accordance with the theorem of § 131, itself normally ordered; and its type defines the sum $\beta_1+\beta_2+\dots+\beta_n+\dots=\beta$. If α_n denotes the sum $\beta_1+\beta_2+\dots+\beta_n$, we see that $\alpha_n=\overline{(M_1, M_2, \dots M_n)}$; and it is clear that $\alpha_{n+1}>\alpha_n$: hence

$$\beta_1=\alpha_1,\quad \beta_2=\alpha_2-\alpha_1,\quad \dots,\quad \beta_n=\alpha_n-\alpha_{n-1}.$$

It will now be shewn (1) that $\beta>\alpha_n$, for every value of n; and, (2) that, if β' is any ordinal number $<\beta$, there is some definite value of n such that α_n $\alpha_{n+1}, \dots$ are all $>\beta'$.

(1) follows from the fact that each α_n is the ordinal number of a segment of $(M_1, M_2, \ldots M_n, \ldots)$ of which β is the ordinal number.

To prove (2), we observe that a segment of $(M_1, M_2, \ldots M_n, \ldots)$ exists, of which β' is the ordinal number, and therefore the element which determines this segment must belong to one of the aggregates $M_1, M_2, \ldots M_n, \ldots$ say M_{n_1}. It follows that the segment is also a segment of $(M_1, M_2, \ldots M_{n_1})$; and therefore, $\beta' < \alpha_{n_1}$, or $\alpha_n > \beta'$, n being $\geqq n_1$.

It has thus been proved that β is the ordinal number which immediately follows all the ordinal numbers $\alpha_1, \alpha_2, \ldots \alpha_n, \ldots$; and it may be spoken of as the limit of the sequence $\alpha_1, \alpha_2, \ldots \alpha_n, \ldots$.

Thus *every ascending sequence* $\alpha_1, \alpha_2, \ldots \alpha_n, \ldots$ *of ordinal numbers determines a limiting number* $\beta = \underset{n=\infty}{L}\ \alpha_n$, *which immediately follows all the numbers of the sequence.*

THE ORDINAL NUMBERS OF THE SECOND CLASS.

137. Every finite ordered aggregate is normally ordered, and its order-type is the ordinal number of the aggregate. The finite ordinal numbers may be spoken of as the ordinal numbers of the first class; to each such ordinal number there corresponds a single cardinal number, and the properties of the finite ordinal numbers are identical with those of the finite cardinal numbers, the terms ordinal and cardinal simply defining the two uses of the same number. In the case of transfinite aggregates there is no such identity between ordinal and cardinal numbers; in fact the arithmetic of the one kind of numbers is essentially different from that of the other kind.

Corresponding to a single transfinite cardinal number there is an infinity of transfinite ordinal numbers; all those transfinite ordinal numbers which correspond to aggregates that have one and the same cardinal number α are said to form a class $Z(\alpha)$, the class of normal order-types which have the cardinal number α.

The ordinal numbers of all those order-types which have the same cardinal number $\aleph_0$ as the aggregate of finite numbers, are said to be of the second class $Z(\aleph_0)$.

The ordinal number $\omega = \underset{n=\infty}{L} . n$, *and is the smallest number of the second class.*

If M denotes the aggregate $(m_1, m_2, \ldots m_n, \ldots)$, then $\bar{M} = \omega$, and $\bar{\omega} = \aleph_0$. Any number β which is $< \omega$, must be the order-type of a segment of M, and M has only segments $(m_1, m_2, \ldots m_n)$ with finite ordinal numbers n; thus β must be a finite number; and therefore the only ordinal numbers $< \omega$ are finite ones.

Every number α *of the second class has a number* $\alpha+1$ *immediately following it.*

For if $\alpha = \overline{M}$, $\bar{\alpha} = \aleph_0$, we have $\alpha + 1 = (\overline{M, e})$, where e is a new element; and since M is a segment of (M, e), we have $\alpha + 1 > \alpha$. Also

$$\overline{\alpha+1} = \bar{\alpha} + 1 = \aleph_0 + 1 = \aleph_0.$$

It has thus been shewn that $\alpha + 1$ is a number of the second class. Every number $< \alpha + 1$, is the order-type of a segment of (M, e); and such segment can only be M, or a segment of M; hence no number $< \alpha + 1$ is $> \alpha$: therefore $\alpha + 1$ is the next number greater than α.

If $\alpha_1, \alpha_2, \ldots \alpha_n, \ldots$ *is any sequence of numbers of the second class, there is a number* $L.\alpha_n$, *also of the second class, which is the smallest number that is greater than every number* α_n *of the sequence.*

If, as in § 136, we write $\beta_1 = \alpha_1$, $\beta_2 = \alpha_2 - \alpha_1 \ldots \beta_n = \alpha_n - \alpha_{n-1}, \ldots$ then if $\overline{G}_n = \beta_n$, we have $L\alpha_n = (\overline{G_1, G_2, \ldots G_n, \ldots})$; and this number $L\alpha_n$ has been shewn to be the smallest number which is $> \alpha_n$ for every value of n. To shew that this number $L\alpha_n$ is of the second class, we have, since $\bar{\beta}_n \leqq \aleph_0$, for every value of n,

$\overline{L.\alpha_n} \leqq \aleph_0 . \aleph_0 \leqq \aleph_0$; and since $\overline{L.\alpha_n}$ is not finite it must therefore $= \aleph_0$.

Two sequences $\{\alpha_n\}$, $\{\alpha'_n\}$ *of numbers of the second class, have the same limiting number, when, and only when, the sequences are related to one another, in accordance with the definition of* § 125.

Let β, γ be the two limiting numbers. and first assume that the sequences are related to one another. If $\beta < \gamma$, then for some value of n, $\alpha'_n > \beta$, $\alpha'_{n+1} > \beta, \ldots$; and hence for some value of n', we must have $\alpha_{n'} > \beta$, $\alpha_{n'+1} > \beta, \ldots$ which is inconsistent with β being the limit of the sequence $\{\alpha_n\}$.

If we assume $\beta = \gamma$, then, since $\alpha_n < \gamma$, for some fixed number r we must have $\alpha'_r > \alpha_n$, $\alpha'_{r+1} > \alpha_n, \ldots$; and similarly, since $\alpha'_n < \beta$, for some fixed number s we must have $\alpha_s > \alpha'_n$, $\alpha_{s+1} > \alpha'_n, \ldots$; hence the two sequences are related to one another.

If n is a finite ordinal number, and α a number of the second class, then $n + \alpha = \alpha$, and hence $\alpha - n = \alpha$.

For $$n + \omega = \omega,$$

since $$n + \omega = \overline{(e_1, e_2, \ldots e_n; f_1, f_2, \ldots f_n, \ldots)}$$
$$= \overline{(g_1, g_2, \ldots g_n, g_{n+1}, \ldots)},$$

where $$g_1 = e_1,\ g_2 = e_2, \ldots g_n = e_n,\ g_{n+1} = f_1,\ g_{n+2} = f_2, \ldots.$$

Further, $$n + \alpha = n + \omega + (\alpha - \omega) = \omega + (\alpha - \omega) = \alpha.$$

If n is a finite number, then $n\omega = \omega$. This is seen, by taking an aggregate of the type ω, and replacing each element by n new elements; then it is clear that the new aggregate is also of type ω.

It can easily be proved that $(\alpha + n)\,\omega = \alpha\omega$, where α is of the second class, and n of the first class.

138. *If α is any number of the second class, then the numbers of the first and second classes, which are less than α, form a normally ordered aggregate of type α, when they are arranged in order as defined above.*

If M is an aggregate such that $\bar{M} = \alpha$, and if α' is an ordinal number $< \alpha$, then there is a segment M' of M such that $M' = \alpha'$; and, conversely, every segment of M determines a number of the first or second class which is $< \alpha$. For, since $\bar{\bar{M}} = \aleph_0$, any segment M' must have either a finite cardinal number, or else must have $\aleph_0$ for its cardinal number. If e_1 is the lowest element of M, a segment M' is determined by an element $e' > e_1$; and every element e', of M, determines a segment M'. If e', e'' are two elements of M, both $> e_1$, and M', M'' the segments of M determined by these elements, and α', α'' their order-types, then if $e' < e''$, it follows by § 132, that $M' < M''$, and hence $\alpha' < \alpha''$.

If then $M = (e_1, M')$, and to the element e' of M', we make the element α' of $\{\alpha'\}$, correspond, the two aggregates M' and $\{\alpha'\}$ are placed in the relation of similarity. It has thus been shewn that $\overline{\{\alpha'\}} = \bar{M}'$; now $\bar{M}' = \alpha - 1 = \alpha$, hence $\overline{\{\alpha'\}} = \alpha$.

Since $\bar{\bar{\alpha}} = \aleph_0$, we have $\overline{\overline{\{\alpha'\}}} = \aleph_0$; and therefore the following theorem is established:—

The aggregate $\{\alpha'\}$ of all those numbers α' of the first and second classes, which are ordinally smaller than a number α of the second class, has the cardinal number $\aleph_0$.

139. *Every number α of the second class is either* (1) *such that it is obtained from a number of the same class immediately preceding it, by the addition of unity, or else,* (2) *such that there exists a sequence $\{\alpha_m\}$ of numbers of the first or second class, having α for its limit.*

Let $\alpha = \bar{M}$; then if M has a highest element e, $M = (M', e)$ where M' is the segment of M determined by e; in this case $\bar{M} = \overline{M'} + 1$, or $\alpha = (\alpha - 1) + 1$.

If M has no highest element, then the aggregate $\{\alpha'\}$ of all numbers $< \alpha$, which is similar to M, has no greatest number; and this aggregate $\{\alpha'\}$ being of cardinal number $\aleph_0$ can be re-arranged as an aggregate $\{\alpha'_n\}$ of type ω. In this aggregate $\{\alpha'_n\}$, some of the numbers α_2', α_3', ... will in general be less than α_1', but others must be greater than α_1'; for α_1' cannot be greater than all the other numbers of the aggregate, there being in $\{\alpha'\}$ no greatest number. Let α'_{p_1} be that number of $\{\alpha'_n\}$ with the smallest index, such that $\alpha'_{p_1} > \alpha_1'$; similarly let α'_{p_2} be that number with the smallest index such that $\alpha'_{p_2} > \alpha'_{p_1}$, and so on. We have now an infinite sequence

$$\alpha_1',\ \alpha'_{p_1},\ \alpha'_{p_2},\ \dots$$

of numbers such that they are in ascending order, and such that their indices are also in ascending order. Since $n \leqq p_n$, we have $\alpha'_n \leqq \alpha'_{p_n}$; hence for every number α' which is less than α there exists a number α'_{p_n} which is $> \alpha'$. Since α is the number which follows next after all the numbers α', it is also the number which follows next after all the numbers $\alpha_1', \alpha'_{p_2}, \alpha'_{p_3}, \ldots$, which we may write as $\alpha_1, \alpha_2, \alpha_3, \ldots \alpha_n, \ldots$; thus $\alpha = L\alpha_n$.

It has thus been shewn that there are two kinds of numbers of the second class, (1) those which have an immediate predecessor in the aggregate of all such numbers arranged in ascending order, and (2) those which have no such immediate predecessor, and are called *limiting numbers*.

A number of the first kind is obtained by means of the first principle of generation, (see § 61), from the immediately preceding number.

A number of the second kind is obtained by the second principle of generation, as the number α which next follows all the numbers α_n of any sequence $\{\alpha_n\}$ of numbers of the second class.

THE CARDINAL NUMBER OF THE SECOND CLASS OF ORDINALS.

140. *The totality of the numbers of the second class arranged in ascending order forms a normally ordered aggregate.*

If A_α denotes the ordered aggregate of all those numbers of the second class which are less than the given number α, then A_α is normally ordered and of type $\alpha - \omega$. For the aggregate $\{\alpha'\}$ of numbers of the first and second classes, which consists of $\{n\}$ and A_α, has been shewn in § 138, to be normally ordered, and thus

$$\{\alpha'\} = (\{n\}, A_\alpha),$$

hence

$$\overline{\{\alpha'\}} = \overline{\{n\}} + \overline{A}_\alpha, \text{ or } \overline{A}_\alpha = \alpha - \omega.$$

Let M denote any part of the aggregate $\{\alpha\}$ of all the numbers of the second class, such that in $\{\alpha\}$ there are numbers which are greater than all the numbers in M; and let α_0 be one such number: then M is a part of A_{α_0+1}, which is such that all the numbers of M are less than at least one number α_0 of A_{α_0+1}. Since A_{α_0+1} is normally ordered, there must be a number α' of A_{α_0+1}, being itself consequently a number of $\{\alpha\}$, which is the next greater number than all the numbers of M. Thus, since $\{\alpha\}$ has a lowest number ω, the conditions are satisfied that $\{\alpha\}$ is a normally ordered aggregate.

It follows by applying the results of § 130, that:—

Every part of the aggregate $\{\alpha\}$ of all numbers of the second class has a least number.

Every such part, in order, is normally ordered.

It will now be shewn that *the aggregate* $\{\alpha\}$ *of all the numbers of the second class, has a cardinal number greater than* $\aleph_0$.

If $\overline{\overline{\{\alpha\}}} = \aleph_0$, the numbers of $\{\alpha\}$ could be arranged in the form

$$\gamma_1, \gamma_2, \dots \gamma_n, \dots$$

of type ω, in which of course the order would not be that of generation. Starting from γ_1, let γ_{p_2} be the γ with the smallest index which is such that $\gamma_{p_2} > \gamma_1$; then let γ_{p_3} be that γ with the smallest index such that $\gamma_{p_3} > \gamma_{p_2}$; and so on. We obtain in this manner a sequence

$$\gamma_1, \gamma_{p_2}, \gamma_{p_3}, \dots$$

in ascending order, the indices $1, p_2, p_3, \dots$ being also in ascending order. In accordance with § 137, there must be a definite number δ of the second class, namely $\delta = L\gamma_{p_n}$, such that $\delta > \gamma_{p_n}$, for every p_n, and consequently such that δ is greater than every γ_n; but this is impossible since $\{\gamma_n\}$ contains every number of the second class; hence $\overline{\overline{\{\alpha\}}}$ cannot equal $\aleph_0$.

Every part of the aggregate $\{\alpha\}$ *of all numbers of the second class has either the cardinal number of* $\{\alpha\}$, *or else the cardinal number* $\aleph_0$, *unless it is a finite part.*

Every such part, when the elements of it are in order of generation, being part of the normally ordered aggregate $\{\alpha\}$, is either similar to $\{\alpha\}$, or else to some segment A_{α_0} of $\{\alpha\}$; hence the cardinal number is either that of $\{\alpha\}$ or is $\overline{\overline{A}}_{\alpha_0} = \overline{\alpha_0 - \omega}$, and this last is either $\aleph_0$, or is finite.

The cardinal number of $\{\alpha\}$ *is the cardinal number next greater than* $\aleph_0$.

If there existed a cardinal number less than $\overline{\overline{\{\alpha\}}}$, and greater than $\aleph_0$, it must be the cardinal number of some part of $\{\alpha\}$; but it has been shewn that every such part of $\{\alpha\}$ has either the cardinal number of $\{\alpha\}$, or is $\aleph_0$, or is finite.

The cardinal number of $\{\alpha\}$, or of $Z\{\aleph_0\}$ is denoted by $\aleph_1$.

THE GENERAL THEORY OF ALEPH-NUMBERS.

141. It has now been shewn that the ordinal numbers of the second class in their order of generation, form a normally ordered aggregate of which the cardinal number is $\aleph_1$, the next greater cardinal number to $\aleph_0$. The ordinal type of the normally ordered aggregate $\{\alpha\}$ of all numbers of the second class, is a number Ω, which is the smallest number of the third class. In analogy with the definition of the second class, and in accordance with what Cantor has denominated the principle of limitation (Hemmungsprinzip), the third class is taken to include all the ordinal types of normally ordered aggregates, of which the cardinal number is $\aleph_1$, and this class is consequently denoted by $Z(\aleph_1)$. The number Ω, which is the order-type of all the numbers of the first and second classes, in the order of generation, and which comes after all those numbers, is not the limiting element of any sequence $\alpha_1, \alpha_2, \ldots \alpha_n, \ldots$ of numbers of the second class; for, as we have seen, every such sequence has a limiting number within the second class. From the point of view adopted by Cantor in his earlier writings, and explained in § 61, in which the successive ordinal numbers are regarded as successively generated, in accordance with postulated principles of generation, the number Ω must be regarded as generated by a third principle of generation, different from the two principles of generation employed in the case of the numbers of the first and second classes. This third principle of generation affirms that every set of ordinal numbers similar to the aggregate of all the numbers of the first and second classes, in their order of generation, is immediately succeeded by a new number, ordinally greater than all the numbers of the set, so that every number which is less than this new number is also less than some of the numbers of the set. When, proceeding from Ω, the numbers $\Omega+1$, $\Omega+2$, ... $\Omega+\Omega$, ... are formed, all three principles of generation will be required, in forming the numbers of the third class.

From the point of view adopted later by Cantor, and explained in the present chapter, Ω is simply defined to be the order-type of the totality of the numbers of the first and second classes, in their normal order. The numbers higher than Ω are then defined in the same manner, each one as the order-type of the totality of the preceding numbers in normal order.

The existence of a whole series of classes of order-types of normally ordered aggregates, *i.e.* of ordinal numbers, has been speculatively asserted by Cantor*, who has however, up to the present time, in his published works, confined his detailed investigations to numbers of the first and second classes.

To each of the successive classes of numbers, there corresponds a single cardinal number, that of the totality of the ordinal numbers up to, and including all the ordinal numbers of that class. The first ordinal number of each class is the order-type of all the numbers of the preceding classes in their order of generation. A new principle of generation is required for the first number of each new class, since that number cannot be regarded as the limiting number of any sequence of which the ordinal number is less than that of the number in question. All the successive principles of generation are however included in the one principle, that an aggregate of normally ordered ordinal numbers has itself an order-type which is a new number; and thus, from this point of view, all the principles of generation, from the second, onwards, are replaced by this one principle.

142. In accordance with this theory, there exists an ordered aggregate

$$1, 2, 3, \ldots n, \ldots \omega, \omega+1, \ldots \omega^2, \ldots \omega^\omega, \ldots \Omega, \Omega+1, \ldots \gamma, \ldots$$

* See *Math. Annalen*, vol. XXI, pp. 587, 588, also vol. XLVI, p. 495.

which contains every ordinal number of every class; and there also exists a similar aggregate

$$1, 2, 3, \dots n, \dots \aleph_0, \aleph_1, \aleph_2, \dots \aleph_n, \dots \aleph_\omega, \dots \aleph_\Omega, \dots \aleph_\gamma, \dots$$

of cardinal numbers, each element of which is the cardinal number of a single class of numbers of the first aggregate.

That the first of these aggregates is normally ordered, may be seen by remarking that if it contained any part, of the type $^*\omega$, then such part would also be part of the normally ordered aggregate formed by the numbers $1, 2, 3, \dots \omega, \dots \alpha$; where α is the highest number in the hypothetical part, of type $^*\omega$. This is impossible, and hence the first aggregate is normally ordered.

Cantor has proved (see § 117) that $\aleph_0$ is less than or equal to the cardinal number of any transfinite aggregate, and that $\aleph_1$ is the cardinal number next greater than $\aleph_0$. A proof has been given by Jourdain†, that $\aleph_2$ is the next greater cardinal number than $\aleph_1$, who has also considered in some detail, the ordinal numbers of the third class, and has given indications of extension to the higher classes.

The question whether every transfinite cardinal number is necessarily an Aleph-number, which is equivalent to asking whether every aggregate is capable of being normally ordered, has engaged a considerable amount of attention. That the answer should be an affirmative one, has been regarded by Cantor as probable. Some discussion of attempts which have been made to settle this matter, will be considered in § 161. A case of great importance is that of the continuum, which is defined as a simply ordered, but not as a normally ordered aggregate. No proof has yet been discovered, of the correctness of Cantor's view, that $c = \aleph_1$. In case c occurs at all in the aggregate of Aleph-numbers, the continuum is capable of being normally ordered. The possibility has also been contemplated that c may be greater than all the Aleph-numbers.

THE ARITHMETIC OF ORDINAL NUMBERS OF THE SECOND CLASS.

143. The ordinal numbers of the second class have been defined as the order-types of normally ordered, enumerably infinite, aggregates; and the operations of addition and multiplication have been defined for these numbers, in §§ 134 and 135. It now remains for us to define exponentials for numbers of this class; and the definition is founded upon the following theorem:—

If ξ is a variable of which the domain consists of the numbers of the first and second classes, including zero, and if γ, δ denote two constants belonging

† See *Phil. Mag.* for 1904, "On the transfinite cardinal numbers of number-classes in general."

to the same domain, such that $\delta > 0$, $\gamma > 1$, then there exists a single-valued determinate function $f(\xi)$, which satisfies the conditions

(1) $f(0) = \delta$.

(2) If ξ', ξ'' are any two values of ξ, such that $\xi' < \xi''$, then $f(\xi') < f(\xi'')$.

(3) For every value of ξ, $f(\xi + 1) = f(\xi) . \gamma$.

(4) If $\{\xi_n\}$ is a sequence of which ξ is the limiting number, then $\{f(\xi_n)\}$ is a sequence of which $f(\xi)$ is the limiting number.

In the case $\delta = 1$, the function $f(\xi)$ is denoted by γ^ξ; and then $f(\xi)$, satisfying the above conditions, defines the exponential function γ^ξ, for all numbers γ, ξ of the first and second classes.

To prove the theorem, in the first place we have

$$f(1) = \delta\gamma,\ f(2) = \delta\gamma^2,\ \ldots,\ f(n) = \delta\gamma^n;$$

thus $f(1) < f(2) < f(3), \ldots$, and the function is determined for every $\xi < \omega$. Next assume that the function is determined for every $\xi < \alpha$, a number of the second class. If α is not a limiting number, $f(\alpha) = f(\alpha - 1)\gamma > f(\alpha - 1)$; and thus $f(\alpha)$ is determined. If α is a limiting number, and is preceded by the sequence $\{\alpha_n\}$, then $\{f(\xi_n)\}$ is a sequence, and $f(\alpha) = Lf(\alpha_n)$. If $\{\alpha_n'\}$ is another sequence such that $\alpha = L\alpha_n'$, then the two sequences $\{f(\alpha_n)\}$, $\{f(\alpha_n')\}$ are related to one another, and therefore have the same limit; and thus $f(\alpha)$ is uniquely determined. $f(\xi)$ is now determined for every ξ: for if there were values of α for which it were not determined, there must be a smallest of such values; the theorem would then hold for $\xi < \alpha$, but not for $\xi \geqq \alpha$; which is contrary to what has been proved above.

144. If α, β are numbers of the first or second class, $\gamma^{\alpha+\beta} = \gamma^\alpha . \gamma^\beta$.

The function $\phi(\xi) = \gamma^{\alpha+\xi}$, satisfies the conditions

(1) $\phi(0) = \gamma^\alpha$.

(2) If $\xi' < \xi''$, $\phi(\xi') < \phi(\xi'')$.

(3) $\phi(\xi + 1) = \phi(\xi) . \gamma$.

(4) If $\{\xi_n\}$ is a sequence such that $L\xi_n = \xi$, then $\phi(\xi) = L\phi(\xi_n)$.

It follows by § 143 that, if we take $\delta = \gamma^\alpha$, then $\phi(\xi) = \gamma^\alpha\gamma^\xi$; hence if $\xi = \beta$, we have

$$\gamma^{\alpha+\beta} = \gamma^\alpha . \gamma^\beta.$$

Again, if α, β are two numbers of the first or second class

$$\gamma^{\alpha\beta} = (\gamma^\alpha)^\beta.$$

If we put $f(\xi) = \gamma^{\alpha\xi}$, we find by applying the theorem of the last section, that $f(\xi) = (\gamma^\alpha)^\xi$, where γ^α replaces γ.

γ being > 2, it can be proved that, for every ξ, $\gamma^\xi \geqq \xi$. The theorem holds for $\xi = 0$, $\xi = 1$; and if it be assumed to hold for all values of ξ which are less than a given number α, then it holds also for $\xi = \alpha$.

For, first let α be not a limiting number: then, if $\alpha - 1 \leqq \gamma^{\alpha-1}$, we have

$$(\alpha - 1)\gamma \leqq \gamma^{\alpha};$$

hence $\gamma^{\alpha} \geqq (\alpha - 1) + (\alpha - 1)(\gamma - 1)$: therefore since $\alpha - 1$ and $\gamma - 1$ are > 1, and $(\alpha - 1) + 1 = \alpha$, we have $\gamma^{\alpha} \geqq \alpha$. If α is a limiting number $= L\alpha_n$, then since $\alpha_n \leqq \gamma^{\alpha_n}$, we have $L\alpha_n \leqq L\gamma^{\alpha_n}$, or $\alpha \leqq \gamma^{\alpha}$. If there were values of ξ such that $\xi > \gamma^{\xi}$, there must be one of such values which is the least of all; and if this were α, then $\xi \leqq \gamma^{\xi}$, if $\xi < \alpha$, but $\alpha > \gamma^{\alpha}$; which is contrary to what has been proved above.

145. Of all the numbers of the second class, the smallest ones are those which are algebraical functions of ω, of the form

$$\omega^n . p_n + \omega^{n-1} . p_{n-1} + \dots + \omega . p_1 + p_0,$$

where $p_0, p_1, \dots p_n$ are finite numbers. If we write

$$\omega_1 = \omega^{\omega},\ \omega_2 = \omega^{\omega_1},\ \omega_3 = \omega^{\omega_2},\ \dots$$

then we obtain the number $\epsilon_0 = L\omega_n$. This number ϵ_0 is the smallest of a species of numbers of the second class which are characterised by the property $\epsilon = \omega^{\epsilon}$, and which Cantor has designated ϵ-numbers. Cantor has shewn that the ϵ-numbers form a normally ordered aggregate of type Ω, and therefore similar to the whole second class of numbers. He has further shewn that every number α of the second class is uniquely representable in the form

$$\alpha = \omega^{\alpha_0}\kappa_0 + \omega^{\alpha_1}\kappa_1 + \dots + \omega^{\alpha_r}\kappa_r,$$

where $\alpha_0, \alpha_1, \alpha_2, \dots \alpha_r$ are numbers of the first or second class which satisfy the conditions $\alpha_0 > \alpha_1 > \alpha_2 \dots > \alpha_r \geqq 0$, and $\kappa_0, \kappa_1, \kappa_2, \dots \kappa_r, r + 1$, are numbers of the first class which are different from zero. For the detailed investigation of the normal form, and for that of the special class of ϵ-numbers, we must refer to Cantor's original discussion*.

THE THEORY OF ORDER-FUNCTIONS.

146. A method of representation of any mode of ordering a given aggregate M has been given by Bernstein†. When the elements of the aggregate are numbers, this method lends itself to a diagrammatic representation of the aggregate as ordered in any particular order-type.

An aggregate M is ordered, in the most general sense of the term, when it is known as regards every pair of elements a, b, whether $a \gtrless b$; but a particular mode of ordering the aggregate can be represented by means of a function $f(a, b)$ of the pairs of elements, which is defined by

$$f(a, b) = 1, \text{ if } a < b;\ f(a, b) = -1, \text{ if } a > b, \text{ and } f(a, a) = 0.$$

* *Math. Annalen*, vol. XLIX, pp. 235—246.

† See his Dissertation, also W. H. Young, on "Closed sets of points and Cantor's numbers," *Proc. Lond. Math. Soc.*, Ser. 2, vol. I.

This function $f(a, b)$ may be denominated an order-function of the aggregate M; and there is one order-function for each possible mode of ordering the aggregate. The function must satisfy the condition that, if $f(a, b)=f(b, c)$, then each equals $f(a, c)$.

Two order-functions $f_1(a, b), f_2(a, b)$ of a given aggregate M, represent two methods of ordering the aggregate in one and the same order-type, provided there exist a reversible transformation ϕ, of the aggregate M into itself, such that $f_1\{\phi(a), \phi(b)\}=f_2(a, b)$.

All those order-functions of a given aggregate M, which correspond to an arrangement of M in one and the same order-type, constitute a family of order-functions; and there is one such family of order-functions corresponding to each order-type in which the given aggregate M can be arranged.

It is clear that the order-functions of a family corresponding to $\bar{M}$ form an aggregate with the same cardinal number as the group of transformations of $\bar{M}$ into itself.

If, in particular, the aggregate M is that of the positive integers, then a pair of elements (a, b) is represented by a cross-point of the rectangular trellis formed in the positive quadrant by drawing all the straight lines, $x=a$, $y=a$, for positive integral values of a, referred to rectangular Cartesian coordinates x, y.

The natural order of the numbers 1, 2, 3, ... will be represented by $f(x, y)$, defined for all the cross-points, so that $f(x, y)=1$, when $x<y$, and $f(x, y)=-1$, when $x>y$, and also $f(x, x)=0$.

Any particular mode* of ordering the numbers 1, 2, 3, ... will be represented by marking one set of cross-points $+1$, and another set -1, those on the diagonal $x=y$, being marked zero.

It is, however, not every mode of so marking the cross-points that represents a possible ordering of the aggregate. That a mode of marking may represent a possible order, two conditions must be satisfied. First, we must have $f(x, y)=-f(y, x)$; and thus points which are optical images relatively to the diagonal $y=x$, must be marked with unities of opposite sign. Secondly, the condition that if $f(a, b_1)=f(b_1, c)$, then each $=f(a, c)$ must be satisfied. This condition may be expressed as follows:—Join every cross-point which is marked $+1$, or 0, with every other such cross-point, then the resulting figure may be called the positive frame-work; join similarly all the pairs of points marked -1, 0; in this way we obtain the negative frame-work. Let lines joining the two pairs of points (x_1, y_1), (x_2, y_2), and (x_1, y_2), (x_2, y_1) be called conjugate lines. The condition which must be satisfied is that no side of the positive frame-work can be conjugate to a side of the negative frame-work. It can easily be seen that the condition so stated is necessary and sufficient.

* Some examples of order-types represented in this manner are given by W. H. Young, *Lond. Math. Soc. Proc.*, Ser. 2, vol. I, p. 244.

THE CARDINAL NUMBER OF THE CONTINUUM.

147. The arithmetic continuum has been defined as an aggregate of the order-type θ (see § 128), and it is thus not normally ordered. It has been held by Cantor* that this aggregate, and perhaps every aggregate, is capable of being arranged as a normally ordered aggregate; but no proof of the correctness of this view has been obtained. If the continuum be capable of arrangement as a normally ordered aggregate, its cardinal number c must be identical with one of the Aleph-numbers; and in fact Cantor has believed that $c = \aleph_1$, the cardinal number of the aggregate of all the order-types of normally ordered enumerable aggregates. As evidence of the probable truth of this view, the facts may be cited that all the sets of points which have actually been defined in connection with the theory of sets of points, have one or other of the two cardinal numbers $\aleph_0$ and c, and that no such set of points has been defined of which it is known that the cardinal number is $> \aleph_0$ and $< c$. This negative evidence is however clearly insufficient to settle the question whether every part of the continuum has one of the powers $\aleph_0$ or c, a question which has hitherto defied all attempts to obtain a conclusive answer. As has already been pointed out, it cannot be assumed that every two cardinal numbers are such as to be comparable with one another; but a proof has been given by G. H. Hardy† that c, and presumably any cardinal number whatever, must either be an Aleph-number, or else be greater than all the Aleph-numbers. The mode of reasoning is a generalization of that employed by Cantor in his proof (see § 117) that $\aleph_0$ is less than any other transfinite cardinal number.

Because $c \geqq \aleph_1$, it is possible to take elements from the number-continuum corresponding to all the numbers of the first and second classes of ordinals. For if this process came to an end, we should have $c = \aleph_0$, which has been proved by Cantor not to be the case. It follows that a set can be selected from the continuum equivalent to the aggregate of ordinal numbers of the first and second classes. Now if a set could be selected from this aggregate equivalent to the continuum, it would follow from the equivalence theorem, proved in § 119, that $c = \aleph_1$; and if no such set could be selected it follows from the definition of inequality in § 112, that $c > \aleph_1$; thus it has been proved that $c \geqq \aleph_1$. If now $c > \aleph_1$, a similar proof would shew that $c \geqq \aleph_2$, and so on. If $c > \aleph_n$ for every finite n, then $c \geqq \aleph_\omega$, and this process may be continued indefinitely through the Aleph series. A criticism of the validity of this proof will be given in § 160.

It is known that every infinite closed linear set of points has one or other of the two cardinal numbers $\aleph_0$, c; and if a set of points can exist of which

* *Math. Annalen*, vol. XXI, p. 550.

† *Quarterly J. of Math.* 1903, p. 87.

the cardinal number has neither of these two values, it must be unclosed, and may without loss of generality be taken as dense-in-itself. The difficulties of dealing with open sets dense-in-themselves are so great, that attempts to find a contradiction involved in the assumption of the existence of such a set, possessing a cardinal number different from both $\aleph_0$ and c, have hitherto been a complete failure.

148. A very remarkable relation has been given by Cantor between the cardinal number of the continuum and that of the integral numbers. This relation is expressed by $c = 2^{\aleph_0}$, or more generally $c = n^{\aleph_0}$, where n is a finite integer.

This theorem was applied by its discoverer to obtain a simple arithmetical proof that the $\aleph_0$-dimensional continuum has the same power as the one-dimensional continuum.

In accordance with the definition of an exponent given in § 116, $2^{\aleph_0}$ is the cardinal number of the proper fractions in the dyad scale,

$$\frac{b_1}{2} + \frac{b_2}{2^2} + \frac{b_3}{2^3} + \dots;$$

where every b is either 0 or 1. In this aggregate each number of the form $\frac{2p+1}{2^q} < 1$, where p and q are integers, occurs twice; hence

$$2^{\aleph_0} = (\overline{\overline{s, X}}),$$

where X is the aggregate of real numbers between 0 and 1, and s is an enumerable aggregate.

It follows from the above, that $2^{\aleph_0} = c + \aleph_0$. Now $c + \aleph_0 = c + 2\aleph_0$, since $\aleph_0 = 2\aleph_0$; therefore $c = c + \aleph_0$; whence we have $2^{\aleph_0} = c$.

From this theorem we deduce $c \,.\, c = 2^{\aleph_0} . 2^{\aleph_0} = 2^{2\aleph_0} = 2^{\aleph_0} = c$; and hence by repeated multiplication by c, we find $c^n = c$, where n is any finite integer.

Again $$c^{\aleph_0} = (2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0 \cdot \aleph_0} = 2^{\aleph_0} = c,$$

and therefore the continuum of finite, or of enumerable dimensions, is equivalent to the one-dimensional continuum.

The aggregate defined by all possible modes of covering the numbers of the continuum by themselves, has the power $c^c \equiv f$, and this number f is greater than c. This has been proved in § 120; for a part of the new aggregate is equivalent to that obtained by replacing the numbers of the continuum either by A or by B, and taking all possible aggregates which arise in this way. Since this part has been shewn to have a cardinal number greater than that of the original aggregate, it follows then that $f > c$.

More generally, if α is any cardinal number, we have $\alpha^\alpha > \alpha$.

If the continuum be divided into any finite number n of parts, such that *all the parts have the same cardinal number, then that cardinal number is the same as that of the continuum**.

The parts may consist of sets of points of any kind.

The theorem may also be stated thus:—if $n\alpha = c$, then $\alpha = c$.

To prove it, we have $n\alpha = c = nc$; and therefore by applying the theorem of § 121, it follows that $\alpha = c$.

149. *The continuum is equivalent to the aggregate of all possible order-types of simply ordered aggregates of cardinal number* $\aleph_0$.

This theorem points the contrast between the aggregate of all order-types of simply ordered enumerable aggregates, which is of power $2^{\aleph_0}$, and the aggregate of all the order-types of normally ordered enumerable aggregates, which is $\aleph_1$. The latter aggregate is, of course, a part of the former one, and thus the theorem $\aleph_1 \leqq 2^{\aleph_0}$, can be deduced.

The theorem may be also stated in the form: *The total number of ways of ordering the integral numbers* 1, 2, 3, ... *is c.*

If μ is the order-type of an enumerable aggregate arranged as a simply ordered aggregate, then a part of an aggregate of the type η, considered in § 127, can always be found which is of the type μ. To establish this, it can be shewn that an aggregate of type μ can always be changed into one of type η, by insertion of new elements. If, between every pair of elements m_1, m_2 of μ, there are other elements, then μ is of one of the types η, $1+\eta$, $\eta+1$, $1+\eta+1$; so that μ is reduced to η by the removal of the lowest and the highest elements, when such elements exist. In any such case, if we add to μ, aggregates of type η, at the beginning and at the end, we obtain an aggregate of type $\eta+\mu+\eta$ which is of type η, whichever of the types η, $1+\eta$, $\eta+1$, $1+\eta+1$ may be identical with μ. If pairs of elements exist in the aggregate of type μ, such that there are no elements between them, an aggregate of type η can be inserted between every such pair, until a new aggregate of one of the types η, $1+\eta$, $\eta+1$, $1+\eta+1$, is obtained; then as before, by adding aggregates of type η, at the beginning and end, we obtain an aggregate of type η. It has thus been shewn that any aggregate of type μ is a part of another aggregate of type η. Since the rational numbers in their totality naturally exist in the order-type η, it follows that an aggregate of any type μ can be made by taking a part of the aggregate of rational numbers, of type η. It follows that the aggregate of all types μ has a cardinal number less than, or equal to, that of the aggregate of all part-aggregates of the set of rational numbers arranged in type η.

Now every aggregate $(r_1, r_2, \ldots)$, of which all the elements are rational numbers, corresponds to a single point of a continuum of an enumerable number

* Bernstein, *loc. cit.*, p. 31.

of dimensions, of which the coordinates are $x_1 = r_1$, $x_2 = r_2$, ... Hence the cardinal number of the aggregate of all part-aggregates of the set of rational numbers is less than, or equal to, the cardinal number of the $\aleph_0$-dimensional continuum, that is, $\leqq c$; and therefore the cardinal number of the aggregate of all types μ is $\leqq c$.

It will now be proved that $c \leqq$ the aggregate of all types μ. To every real number between 0 and 1, there corresponds an infinite sequence $b_1 b_2 b_3 \ldots$ where every b is either 0 or 1, expressing the number in the dyad scale. After each b, insert an aggregate of type π, and we then have an aggregate $b_1 \pi b_2 \pi b_3 \pi \ldots$, of type $\nu = b_1 + \pi + b_2 + \pi + b_3 + \pi + \ldots$. Here, some of the b's may be zero, and these may be simply omitted; thus $\pi + 0 + \pi = \pi + \pi$. Hence to any real number x between 0 and 1, there corresponds the type

$$\nu = b_1 + \pi + b_2 + \ldots .$$

It is now necessary to shew that the two order-types ν, ν', which correspond to two different numbers, x, x', are necessarily distinct from one another. If $\nu = \nu'$, we can write the equality $C_1 + \pi + \zeta_1 = C_1' + \pi + \zeta_1'$, where C_1, C_1' are each either 0 or 1; and from this we obtain, by means of the theorem of § 129, $C_1 = C_1'$, and $\zeta_1 = \zeta_1'$. The last equation can be written

$$C_2 + \pi + \zeta_2 = C_2' + \pi + \zeta_2',$$

and from this we conclude that $C_2 = C_2'$, $\zeta_2 = \zeta_2'$; and we can proceed onwards in the same manner. From $b_1 = b_1'$, $b_2 = b_2'$, ..., we conclude that $x = x'$. It has thus been shewn that $\{x\} \simeq \{\nu\}$; and from this, we conclude that $c \leqq$ the cardinal number of all order-types μ.

This part of the theorem is due to Cantor*, and the first part to Bernstein. By combining the two results, the complete theorem is established.

This important result may also be expressed by saying that *the totality of all permutations of the sequence of positive integers has the power of the continuum.*

It may also be shewn that *the totality of all parts of the sequence* 1, 2, 3, ... *has the power of the continuum.*

For if we form a sequence by writing 0 for each of the numbers 1, 2, 3, ... which does not occur in a given part of (1, 2, 3, ...); and 1 for each number which does occur in the given part, then the sequence of 0's and 1's thus obtained, corresponds to a real number expressed in the dyad scale, and therefore the numbers of the continuum are put into correspondence with the parts of the sequence (1, 2, 3, ...).

150. It can be shewn that *the aggregate of all sets of points in the n-dimensional continuum has a cardinal number greater than c.*

* See Bernstein's Dissertation, p. 7.

For in the aggregate $\{P\}$ of all points in an n-dimensional continuum, we can substitute 0 for each point P which does not occur in a given set of points of the continuum, and 1 for each point P which does occur; we then obtain an aggregate consisting of 0's and 1's: but it is known that the totality of all such aggregates has the power 2^c, which is $>c$.

On the other hand, *the totality of all closed sets of points in the n-dimensional continuum has the same power c as the continuum.*

Every closed set is the derivative of an enumerable set of points; and to every enumerable set of points there corresponds a single closed set.

It follows that the cardinal number of the totality of closed sets is $\leqq$ the cardinal number of the totality of enumerable sets of points chosen out of the continuum. To shew that the latter is c, we observe that it is $\leqq$ the aggregate of all combinations of points of the continuum in sets of $\aleph_0$ elements, that is, $\leqq c^{\aleph_0}$, or $\leqq c$.

Again, every single point of the $\aleph_0$-dimensional continuum corresponds to a single point of the one-dimensional continuum, and this point is an enumerable part of the continuum; hence the totality of enumerable sets of points of the n-dimensional continuum is $\geqq c$. On combining this with what has been proved above, we see that the totality of all enumerable sets of points in the n-dimensional continuum is c; hence the totality of all closed sets of points in the n-dimensional continuum is $\leqq c$.

Again, the totality of all closed sets of points in the n-dimensional continuum is $\geqq c$. For one such closed set can be taken in each of an infinity of the domains $x_1 = \alpha$, where x_1 is one of the n coordinates which determine the position of a point in the n-dimensional continuum; and the aggregate of all possible values of α has the power c. We thus obtain an aggregate of closed sets which has the power c; and it follows that the aggregate of all closed sets in the n-dimensional continuum is $\geqq c$.

Since the totality of all closed sets of points in the n-dimensional continuum is $\geqq c$, and at the same time is $\leqq c$, it must have the cardinal number c.

Since* every curve or surface in a continuum is formed by a closed set of points, we see that every possible curve or surface corresponds uniquely to a single definite real number.

151. A method of constructing a set of points of which the cardinal number is $\aleph_1$, has been given† by G. H. Hardy.

If we start from the sequence

$$1, 2, 3, 4, 5, \ldots\ldots (1)$$

* Bernstein, *loc. cit.*, p. 43.

† *Quarterly Journal of Math.*, vol. xxxv, 1903, "A theorem concerning the infinite cardinal numbers." A criticism of this construction will be given in § 162.

of integral numbers, a new sequence

$$2,\ 3,\ 4,\ 5,\ \ldots\ldots\ (2)$$

is formed by omitting the first term.

Continuing this process, we form

$$3,\ 4,\ 5,\ 6,\ \ldots\ldots\ (3)$$
$$4,\ 5,\ 6,\ 7,\ \ldots\ldots\ (4)$$
$$5,\ 6,\ 7,\ 8,\ \ldots\ldots\ (5)$$
$$\ldots\ldots\ldots$$

We now form a new sequence

$$1,\ 3,\ 5,\ 7,\ 9,\ \ldots\ldots\ (\omega)$$

by traversing the above infinite array of sequences diagonally. Then we form

$$3,\ 5,\ 7,\ 9,\ 11,\ \ldots\ldots\ (\omega+1)$$
$$5,\ 7,\ 9,\ 11,\ 13,\ \ldots\ldots\ (\omega+2)$$
$$7,\ 9,\ 11,\ 13,\ 15,\ \ldots\ldots\ (\omega+3)$$
$$9,\ 11,\ 13,\ 15,\ 17,\ \ldots\ldots\ (\omega+4)$$
$$\ldots\ldots\ldots$$
$$1,\ 5,\ 9,\ 13,\ 17,\ \ldots\ldots\ (\omega.2)$$
$$5,\ 9,\ 13,\ 17,\ 21,\ \ldots\ldots\ (\omega.2+1)$$
$$9,\ 13,\ 17,\ 21,\ 25,\ \ldots\ldots\ (\omega.2+3)$$
$$\ldots\ldots\ldots$$
$$1,\ 9,\ 17,\ 25,\ 33,\ \ldots\ldots\ (\omega.3)$$
$$\ldots\ldots\ldots$$

Thus sequences corresponding to all the numbers $\omega.\mu+\nu$ can be formed.

To form the sequences corresponding to ω^2, we take the array of sequences

$$1,\ 3,\ 5,\ 7,\ 9,\ \ldots\ldots\ (\omega)$$
$$1,\ 5,\ 9,\ 13,\ 17,\ \ldots\ldots\ (\omega.2)$$
$$1,\ 9,\ 17,\ 25,\ 33,\ \ldots\ldots\ (\omega.3)$$
$$1,\ 17,\ 33,\ 49,\ 65,\ \ldots\ldots\ (\omega.4)$$
$$1,\ 33,\ 65,\ 97,\ 129,\ \ldots\ldots\ (\omega.5)$$
$$\ldots\ldots\ldots$$

and traverse it diagonally; we thus obtain

$$1,\ 5,\ 17,\ 49,\ 129,\ \ldots\ldots\ (\omega^2).$$

Generally, if $b_1, b_2, b_3, b_4, \ldots$ is the sequence corresponding to β, the sequence $b_2, b_3, b_4, b_5, \ldots$ corresponds to $\beta+1$. To obtain a sequence corresponding to a number γ which is a limiting number of the second class, we take the array of sequences corresponding to any ascending set of numbers $\beta_1, \beta_2, \ldots$ of which the limit is γ, and traverse it diagonally. It is clear that, in this manner, a sequence can be found for any given number of the second class; but that the set of sequences so obtained is not unique. For example, ω might have been taken as the limit of 1, 3, 5, 7, ..., or ω^2 might have been taken as the limit of $\omega+1$, $\omega.2+2$, $\omega.3+3$, ...

It will be shewn that the sequences $b_1, b_2, b_3, \ldots$ can be so chosen that in every case $b_1 < b_2 < b_3, \ldots$; and that, if the sequences $b_1, b_2, \ldots$ and $b_1', b_2', \ldots$ correspond to any two numbers β, β', where $\beta < \beta'$, then there exists a number N such that $b_n' > b_n$, for $n \geqq N$; and thus that the sequences are distinct from one another.

Let us assume that sequences, corresponding to all numbers $< \gamma$, have been constructed in such a manner that this condition is satisfied. First, let γ be a non-limiting number, so that $\gamma = \gamma' + 1$. Then if $\beta < \gamma'$, there is a number N such that $a_n' > b_n$, for $n \geqq N$, where $a_1', a_2', a_3', \ldots$ is the sequence which corresponds to γ'. But if $a_1, a_2, a_3, \ldots$ is the sequence which corresponds to γ, we have $a_n = a'_{n+1} > a_n' > b_n$, for $n \geqq N$.

Hence, if the construction is possible for all numbers $< \gamma$, it is possible for all numbers $\leqq \gamma$, where γ is a non-limiting number.

Next, let us suppose that γ has no immediate predecessor, and that $\gamma = L.\beta_m$; then also $\gamma = L(\beta_m + \nu_m)$, where the ν_m are finite numbers. Now there is a number N_1, such that $b_{2,n} > b_{1,n}$, for $n \geqq N_1$, where $b_{m,n}$ denotes the nth number in the sequence corresponding to β_m. *A fortiori*, if $\gamma_m = \beta_m + \nu_m$, we have $c_{2,n} = b_{2,n+\nu_2} > b_{2,n} > b_{1,n}$, for $n \geqq N_1$, where $c_{m,n}$ is the nth number in the sequence corresponding to γ_m. But if we take $\nu_2 > b_{1,N_1-1}$, we have $c_{2,n} = b_{2,n+\nu_2} \geqq n + \nu_2 > b_{1,N_1-1} > b_{1,n}$ for $n < N_1$, and hence we have $c_{2,n} > b_{1,n}$, for all values of n. Similarly ν_3 can be so chosen that $\gamma_3 > \gamma_2$, and $c_{3,n} > c_{2,n}$, for all values of n; and so on generally. If we write γ_1 for β_1, and $c_{1,n}$ for $b_{1,n}$, we have a doubly infinite array

$$\begin{array}{llll} c_{1,1}, & c_{1,2}, & c_{1,3}, & \ldots \\ c_{2,1}, & c_{2,2}, & c_{2,3}, & \ldots \\ c_{3,1}, & c_{3,2}, & c_{3,3}, & \ldots \end{array}$$

and we define the sequence corresponding to γ, by traversing it diagonally, so that $c_n = c_{n,n}$. If then $\beta < \gamma$, we can find m so that $\beta < \gamma_m$; then there is a number K, such that $c_{m,n} > b_n$, for $n \geqq K$. But if $n > m$, we have $c_n = c_{n,n} > c_{m,n}$; and thus if n is greater than the greater of the two numbers m, K, we have $c_n > b_n$. It has thus been shewn that if the construction is possible for all numbers $< \gamma$, it is possible for all numbers $\leqq \gamma$, whether γ is a limiting number or not.

In this manner a sequence is obtained which corresponds to any assigned number γ of the second class, and this sequence is distinct from those which correspond to the numbers $< \gamma$, such sequences being also distinct from one another.

The sequences may be correlated with points in the linear continuum (0, 1). To correlate a sequence $b_1, b_2, b_3, \ldots$, we may take the binary radix fraction in which the b_1th, b_2th, b_3th, ... figures are all 1, and the remaining figures all 0. In this manner a set of points is shewn to exist, such that one point of the set corresponds to each number of the first or of the second class. This amounts to the construction of a set of points of cardinal number $\aleph_1$. Just as an enumerable set of points is determinate, when the point which corresponds to any assigned number n of the first class is determinate, so the set of cardinal number $\aleph_1$ is determinate, in the sense that a definite point is determined corresponding to any assigned number β of the first or of the second class.

It may be remarked that *a set of points of cardinal number $\aleph_1$, or of any cardinal number $> \aleph_0$, when arranged in normal order, cannot possibly be in the order in which they occur in the continuum.*

For if a set of points, in the order in which they occur in the continuum, forms a normally ordered aggregate, each point and the next succeeding one define a linear interval of which they are the end-points. We have thus a set of intervals which must have the same cardinal number as the given set of points. Each interval of the set abuts on the next one, and thus the end-points together with their limiting points define an enumerable closed set. Hence the given set must be enumerable.

GENERAL DISCUSSION OF THE THEORY.

152. An account having now been given of the abstract theory of aggregates, as developed by Cantor and others, the remainder of this chapter will be devoted to a critical discussion* of the theory.

In accordance with Cantor's general theory of ordinal numbers, and of aleph-numbers, there exist two aggregates

$$1, \;\; 2, \ldots n, \;\; \ldots \omega, \;\; \omega + 1, \ldots \;\Omega, \Omega + 1, \ldots \beta, \ldots,$$

$$\aleph_0, \aleph_1, \ldots \aleph_n, \ldots \aleph_\omega, \aleph_{\omega+1}, \ldots \;\aleph_\Omega, \aleph_{\Omega+1}, \ldots \aleph_\beta, \ldots,$$

the first, the aggregate of all ordinal numbers, and the second that of all $\aleph$ cardinal numbers. These aggregates are both normally ordered, and are similar to one another; and they contain, respectively, every ordinal number, and every cardinal number which belongs to a normally ordered aggregate.

In accordance with the principle which is fundamental in the whole theory, that every normally ordered aggregate has a definite order-type, which is its ordinal number, and has also a definite cardinal number, it is seen that the above aggregates have an ordinal number γ, and a cardinal

* Most of the critical remarks here made have been published in the *Proc. Lond. Math. Soc.*, ser. 2, vol. III.

number $\aleph_\gamma$. The ordinal number γ must itself occur in the first aggregate, and must therefore be the greatest ordinal number, *i.e.* the last element of the aggregate; moreover $\aleph_\gamma$ must occur in the second aggregate, and must be the last element of that aggregate. There can, however, be no last ordinal number; for, on the assumption of the existence of γ, an aggregate of ordinal number $\gamma+1$, can be formed. For example, by placing the first element of either of the above aggregates after γ or after $\aleph_\gamma$ respectively; it can at once be shewn that there is no last ordinal number, and consequently no last aleph-number $\aleph_\gamma$. We have thus arrived at a contradiction.

Burali-Forti, who first pointed out this contradiction†, accounted for it by denying the truth of the theorem, that any two distinct ordinal numbers α_1, α_2 must necessarily satisfy one of the relations $\alpha_1 > \alpha_2$, $\alpha_1 < \alpha_2$, in accordance with the definition which has been given in § 134, of the meaning of these relations. However, Cantor's proof of this theorem (see § 133), does not appear to be capable of refutation, and consequently the origin of the contradiction cannot be explained in the manner indicated.

B. Russell has suggested‡ that the aggregates of all ordinal numbers and of all aleph-numbers are not normally ordered, and therefore that these aggregates have no ordinal number, and that their cardinal number is consequently not necessarily an aleph-number. He admits, however, that the segments of either aggregate are normally ordered. This explanation is confuted by the argument that, if the above aggregates are not normally ordered, then they must contain parts, of type $^*\omega$; such a part would then be a part also of a segment of one of the aggregates, and such segment would not be normally ordered.

The contradiction has been explained by Jourdain§, by means of the suggestion that there are ordered aggregates which have no order-type, and no cardinal number; and that the above aggregates belong to such class. To such aggregates he gives the name *inconsistent* aggregates, in virtue of the fact that, of such an aggregate it is impossible to think, without contradiction, as a "collection by the mind of definite and distinct objects *to a whole*." It appears from a statement made by Jourdain ‖, that Cantor had himself, some years previously, arrived at the same conception and name.

In accordance with this view of the matter, there exists an ordered aggregate, viz. that of all the ordinal numbers, every segment of which is normally ordered, and has a cardinal number, and yet such that the aggregate

† *Rend. del. circolo mat. di Palermo*, vol. XI, 1897, "Una questione sui numeri transfiniti."

‡ *The Principles of Mathematics*, vol. I, p. 323.

§ *Phil. Mag.* 1904, "On the transfinite numbers of well-ordered aggregates."

‖ *Loc. cit.* p. 67, note; see also Hilbert, *Jahresbericht der deutsch. math. Vereinig.* vol. VIII, p. 184.

itself, being "inconsistent," cannot, without contradiction, be thought of as háving a definite order-type. This amounts to a denial of the universal validity of the fundamental principle that every ordered aggregate has a definite order-type; and yet it is by means of this very principle that the existence of the successive ordinal numbers is regarded as having been established. Each successive ordinal number was defined to *be* the order-type of the ordered aggregate of all the preceding ordinal numbers. The doubt thus thrown upon the validity of the principle by means of which the existence of the complete series of ordinal numbers, and simultaneously, that of the aleph-numbers, is established in Cantor's theory, naturally suggests that a further scrutiny of the foundations of that theory is required. It is not clear, *à priori*, that an aggregate which is inconsistent, in the sense employed above, may not be reached at an earlier stage of the process of forming the successive classes of ordinal numbers, before the aggregate of all such numbers, in the sense of Cantor's theory, is reached. Moreover, it would seem reasonable to expect, that so fundamental a distinction, as that involved in the notion of an inconsistent aggregate, should be indicated in the general definition of an ordered aggregate, or in close connection therewith. In any case, an explanation of the contradiction, on these lines, cannot be regarded as satisfactory, until criteria have been obtained which shall suffice to decide, in respect of any particular ordered aggregate, whether such aggregate has an order-type and a cardinal number, or whether it is an inconsistent aggregate.

153. Before proceeding to attempt the consideration of how far Cantor's general theory of ordinal numbers and aleph-numbers can be accepted, we shall examine the definition of an aggregate in general, with a view to discovering whether it has, in the form given in § 111, the requisite degree of precision. An attempt will then be made to decide what limitations or qualifications must be imposed upon the nature of an aggregate, so that, in the development of the theory, the possibility of being confronted by such a contradiction as that which was pointed out by Burali-Forti, may be removed at its source.

The term aggregate being taken as denoting a collection of distinct objects, in the most general sense, the difficult question arises as to the conditions under which the elements that form the aggregate can be regarded as adequately defined. In the case of a finite aggregate, the elements may be defined by means of individual specification, but this is not possible in the case of a transfinite aggregate; individual specification must then, in the latter case, be replaced by a law, or by a finite set of laws, forming the *norm* by which the aggregate is defined. *Primâ facie* the most general definition of an aggregate which presents itself is that an aggregate consists of all objects, such that each satisfies certain specified conditions. It is however convenient to admit the case of two or more alternative sets of conditions; thus an aggregate may contain all objects, each of which satisfies either the conditions (A), or else one of the sets of conditions (B), ... (K). The conditions forming the norm by which the aggregate is defined must be of a sufficiently precise character to make it logically determinate, as regards any particular object whatever, whether such object does or does not belong to the aggregate. As we have seen, for example, in the case of the aggregate of algebraical numbers, the means at our disposal may not suffice to render the actual determination possible, in any particular case; we therefore agree to fall back upon the logical determinacy as sufficient; thus it is logically determinate as regards a number, defined in any particular manner, whether that number is algebraic or not, and consequently we regard all the algebraical numbers as forming an aggregate in accordance with the definition of that notion. We shall accordingly define the term aggregate, as follows:—

All objects which are such as to satisfy a prescribed norm are said to belong to an aggregate defined by that norm. The norm consists of a set of specified conditions, or of a specified set of alternative specified conditions; and this norm must be sufficient to render it logically determinate, as regards any particular object whatever, whether that object belongs to the aggregate or not.

It is clear that the elements of an aggregate, being subject to a common norm, must have a certain community of nature which constitutes the ground of the aggregation.

In the case of a finite aggregate, the norm may take the form of individual specification of the objects which form the aggregate.

154. It is not clear that an aggregate defined in the above sense is necessarily capable of being ordered at all. For example, it is difficult to see that such an aggregate as that of "all propositions" could conceivably be ordered; where it is assumed that the meaning of the word "proposition" is taken as so definite, that this aggregate has a norm in accordance with the definition above. Again, to take an example among aggregates of the kind usually considered in Mathematical theory, we may consider the aggregate obtained by covering the aggregate of real numbers by itself. This aggregate which has the cardinal number $f \equiv c^c$, is equivalent to the aggregate of all the functions of a real variable; it is difficult, if not impossible, to see how order could be imposed upon this aggregate. If then, a transfinite aggregate is to be given as an ordered aggregate, or is to have an order imposed upon it, or rather discovered in it, it would appear to be necessary that the norm, which constitutes the definition of the aggregate, should be of such a character, that a principle of order is contained therein, or can at all events be adjoined thereto; so that, when any two particular elements are considered, the conditions which they satisfy in virtue of their belonging to the aggregate, when individualized for the particular elements, may be sufficient also to allow of relative rank being assigned to those elements in accordance with a principle of order. This is in fact the case in such aggregates as those of the integral numbers, the rational numbers, or the real numbers. In the case, for example, of the positive rational numbers, the relative rank of any two particular elements (p, q), (p', q') is assigned by the system of postulations, contained in § 11, which defines the aggregate. It may, of course, also be possible in other cases, as in this one, to re-order the aggregate, in accordance with some other law, extrinsically imposed upon the aggregate; but the nature of the elements must be such that this is possible.

We can now state that:—

In order that a transfinite aggregate defined as in § 153, *may be capable of being ordered, a principle of order must be explicitly or implicitly contained in the norm by which the aggregate is defined.*

The relative order of any two elements chosen from an ordered aggregate depends upon the individual characteristics of those elements, in accordance with the principle of order.

In the definition of order-type given by Cantor (see § 122), according to which the order-type of an aggregate is obtained by making abstraction of the particular nature of the elements of the aggregate, it is assumed that the aggregate is given as an ordered aggregate. Again, in his definition of cardinal number (see § 111), Cantor has assumed that the aggregate is given as an ordered one; the cardinal number there appears as the result of a double abstraction, viz. of the particular nature of the elements, and of the order in which they are given. The question however arises, whether the definition of cardinal number should not be such as to be also applicable in the case of aggregates which are not given as ordered aggregates. Cantor has himself, in fact, in his theory of exponentials involving transfinite cardinal numbers, contemplated certain aggregates as having cardinal numbers, whilst such aggregates were not given as ordered aggregates, and *primâ facie*, at all events, are not capable of being ordered.

155. Taking the case of an aggregate defined as an ordered aggregate, we now approach the consideration of the fundamental question, whether, and under what conditions, if any, such an aggregate can be regarded as having a definite order-type, and a definite cardinal number. This is equivalent to asking whether, or when, meanings can be given to those terms, in accordance with general definitions, of such a character that they can be treated as permanent objects for thought, or as mathematical entities which may themselves be elements in aggregates.

With reference to Cantor's definition (see § 111) of the cardinal number of a transfinite aggregate, by abstraction, in accordance with which the cardinal number is represented by replacing each element by an abstract unity, it must be observed that such a substitution would replace the given aggregate by another one which had no longer any intelligible relation with the norm by which the original aggregate is defined. The abstract unities would be indistinguishable from one another, and the new aggregate would be indistinguishable from any other non-finite aggregate of such *unities*. It would be impossible to decide, as regards any particular abstract unity, whether it belonged to the aggregate or not; in fact, to make complete abstraction of the individual nature of the elements of an aggregate is to destroy the aggregate. A definition by abstraction could be justified only by the interpretation, that abstraction is made of those characteristics only, in which the elements of the aggregate differ from the corresponding elements of all possible equivalent aggregates. Thus the existence of aggregates equivalent to the given aggregate would appear to be essential, if the latter is to be regarded as having a cardinal number to which any definite meaning can be attached. On the grounds stated, the definition of a cardinal number, as the characteristic or class-name, of a class of equivalent aggregates, is to be preferred to the definition given by Cantor. Accordingly, an aggregate has a cardinal number, only when it is one of a plurality of equivalent aggregates distinct from one another. The elements of one of these aggregates must be essentially different from those of another of them; it would not, for example, be admissible to consider two equivalent normally ordered aggregates as essentially different from one another, when the one can be obtained from the other by replacing the elements of some segment by other elements, the remainder being left unaltered. In all cases the correspondence between equivalent aggregates must be definable by some norm.

We are thus led to the following statement containing a definition of cardinal number:—

The members of any particular class of equivalent aggregates have a quality in common in virtue of their equivalence. The name of this quality of mutual equivalence is the cardinal number, and may be regarded as characteristic of each aggregate of the particular class.

In Cantor's definition of the order-type of a simply ordered transfinite aggregate (see § 122), abstraction is made of the nature of the elements, their order in the aggregate being alone retained. The order-type is then regarded* as represented by an aggregate of abstract unities, in the order of the elements of the given aggregate. In any ordered aggregate, it is however the individual characteristics of any two elements which determine their relative order in the aggregate, in accordance with some principle of order valid for the whole aggregate. If complete abstraction be made of the characteristics of the various elements, order has then disappeared from the aggregate. It must be supposed, that in Cantor's representation of the order-type, there are attached to the abstract unities marks of some kind,

* See *Math. Annalen*, vol. XLVI, p. 497.

which may in particular cases be marks indicating position in space or time, by which the order of the various abstract *unities* is denoted; the given aggregate is then really replaced by an aggregate of these marks, and the abstract *unities* are superfluous. These marks, by which order is determined, must also have been associated with the elements of the original aggregate. It thus appears, that in a definition by abstraction, it can be only those characteristics (if any) of the various elements which are irrelevant in determining the order, of which abstraction is made: thus the aggregate is really replaced by a similar one. On these grounds, that definition of an order-type is to be preferred, in which the order-type is defined as the characteristic, or class-name, of a class of similar aggregates. Accordingly, in order that a given aggregate may have an order-type, to which a definite meaning can be attached, it is necessary that the aggregate be one of a plurality of similar aggregates.

We may accordingly state that:—

The members of any particular class of similar aggregates have a quality in common, in virtue of their relation of similarity. This quality of mutual similarity possessed by the aggregates is their order-type, and may be represented by a name or symbol, regarded as characteristic of each aggregate of the particular class.

The considerations above adduced may be applied in the case of an aggregate which is a segment of the hypothetical aggregate of all ordinal numbers. In this case it is impossible to make abstraction of the nature of the individual elements of the aggregate, without destroying the order, because the elements are themselves nothing more than marks indicating order. Hence it would appear, that the aggregate cannot in any intelligible sense be considered as having an order-type, unless it be possible to define an aggregate of objects of some other kind, which shall be similar to the one under consideration.

156. We proceed to consider, from a somewhat different point of view, those aggregates which consist of ordinal numbers in their order of generation.

There are two distinct methods of establishing the existence of a class of mathematical entities.

(1) Their existence, as definite objects for thought, may be shewn to follow as a logical consequence of the existence of other entities already recognized as existent, or of principles already recognized as valid; so that the existence of the new entities in question cannot be denied without coming into contradiction with truths already known. This method may be termed the *genetic method.*

(2) The existence of the entities may be postulated; and their mutual relations, and their relations with other entities already known to exist, may be defined by means of a complete system of definitions and postulations.

Accordingly, the objects in question are a relatively free creation of our mental activity. The validity of the scheme thus set up is established when it is shewn to be free from internal contradiction. Its utility is to be judged by its applicability to the general purposes of the science, and by the light it may throw upon the fundamental principles of that science, in virtue of the scheme containing a generalization of what was previously known. This method may be termed the *method of postulation*. It may, however, be urged that the failure to discover contradictions within a scheme which has been postulated is no proof that such contradictions do not exist, and that such proof can only be supplied by the exhibition of a system of entities already known to exist, such that the relations between them are in accordance with those postulated in the scheme in question.

Both these methods have been employed by Cantor in his theory of transfinite numbers and order-types. In his earlier treatment of the subject, he employed the second of the above methods. The existence of the new number ω, and of the limiting numbers of the second class, was postulated, in accordance with the second principle of generation. Freedom from contradiction, and utility in connection with the theory of sets of points, which suggested the postulations, were relied upon as the grounds upon which the system of new numbers was to be justified. The first number Ω, of the third class, was introduced by a new postulation.

In his later and more abstract treatment of the subject, an account of which has been given in the present chapter, Cantor applied the genetic method. The existence of the number ω is not directly postulated, but is taken to follow from the existence of the aggregate $\{n\}$, of integral numbers; ω is defined to be the order-type of this aggregate, and it is assumed that such order-type is a definite object which can itself be an element of an aggregate. The existence, as definite entities, of the cardinal numbers being assumed, the successive ordinal numbers of the successive classes are obtained by assuming as a general principle, that an ordered aggregate necessarily possesses a definite order-type which can be regarded as itself an object, the ordinal number coming immediately after all those that are the elements of the aggregate of which it is the order-type.

It has been seen above, that the assumptions that an ordered aggregate necessarily possesses a definite order-type, and that it also possesses a definite cardinal number, both of which can be regarded as objects, lead to the contradiction pointed out by Burali-Forti. It appears, therefore, that the class of entities, which is constituted by the ordinal numbers of all classes, and the similar aggregate of aleph-numbers, do not satisfy the condition of being subject to a scheme of relations which is free from contradiction. In fact, the principle, in accordance with which their existence is inferred, conflicts with the definition of the aggregates as containing respectively every ordinal number, and every aleph-number. It would then appear, that the genetic process which led to the definition of the aggregates of all ordinal numbers, and of all aleph-numbers, cannot be a valid one. Thus the principle that every ordered aggregate has a definite order-type, which may be regarded as a permanent object of thought, cannot be accepted as a universal principle to be used in a genetic mode of establishment of the existence of a class of entities. A denial of the validity of this principle does not however preclude the less ambitious procedure of postulating the existence of definite ordinal numbers of a limited number of classes, in accordance with Cantor's earlier method. So long as the postulation of the existence of ordinal numbers does not go beyond some definite point, no contradiction will arise, and the validity of the scheme, for purposes of representation, will suffice to justify the postulations which are made. An attempt to examine the structure of such a class of ordinal numbers, as that of the ωth class, with cardinal number $\aleph_\omega$, or that of the Ωth class, with cardinal number $\aleph_\Omega$, will lead to the conviction that such conceptions are unlikely to prove capable of useful application in any branch of Analysis or of Geometry. Nevertheless, should inexorable logic compel us to contemplate the existence of such classes of objects, they would be a proper field of exploration; we have however seen that there are grave doubts as to whether this be in fact the case.

157. The genetic method being rejected on the ground that it leads to the construction of a class of entities which in its entirety can have no existence, we have to fall back upon the method of postulation. A consideration of the essential elements in the conceptions which lie at the base of the scheme of finite integral numbers may afford guidance as to how far we may properly proceed in the construction, by postulation, of transfinite ordinal numbers of successive classes. The ordinal numbers of any one particular class are those which belong to rearrangements of the elements of an aggregate, of which aggregate the order-type is the lowest number of that class. We may therefore consider primarily, the lowest numbers of the classes of which the cardinal numbers are $\aleph_0$, $\aleph_1$, $\aleph_2$ ... respectively. It was pointed out in Chapter I., in the case of the finite numbers, that the existence of an integral number does not follow as a mere logical consequence of the existence of the preceding numbers, but that each ordinal, or each cardinal number appears as the characteristic of the members of a family of similar, or of equivalent, aggregates of objects, the number in question being then the ordinal, or the cardinal, number of each member of the family. Thus the notion of correspondence between the elements of different aggregates was seen to be an essential element in the conception of either an ordinal, or a cardinal, number as characteristic of a class of aggregates. In the genetic method, as applied to the construction of the whole series of classes of transfinite ordinal numbers, this notion of correspondence between the elements of different aggregates having the same number plays no part; and in fact, the existence of a number is constantly inferred from that of a single unique ordered aggregate. For example, the existence of Ω, and of $\aleph_1$, is inferred from the existence of the single aggregate of numbers of the first and second classes. Generally, in the whole scheme, the existence of a new number is inferred from the existence of that unique aggregate which contains the preceding ordinal numbers. That this procedure leads to contradiction has been already seen. The transfinite numbers must be regarded as obtained, or defined, in accordance with the same principles as hold good in the case of the finite numbers, if they are to be regarded as numbers, even in an extended sense of that term. It seems then highly probable, that the neglect of the principle, that correspondence between similar or between equivalent aggregates is essential to our right to consider the numbers belonging to aggregates as definite entities, may be the source of the contradiction which arises from the thoroughgoing application of the genetic method that leads to Cantor's complete series of ordinal numbers and aleph-numbers. In accordance with this view of the nature of Number, finite or transfinite, the postulation of the existence of a definite entity, which entity shall be entitled to be regarded as a number, is only justified when it is shewn that other aggregates exist besides the aggregate which consists of the preceding ordinal numbers, of which other aggregates the postulated number is either the characteristic ordinal or the cardinal number. Thus the postulation of the existence of the numbers ω, and $\aleph_0$, requires for its justification, the exhibition of other aggregates besides $\{n\}$, that of all finite numbers; in this case the requirement is satisfied by the definition of sets of points, or of other geometrical objects, and thus there really exists a class of aggregates which is similar to the ordered aggregate $1, 2, 3, \ldots n \ldots$; and hence the postulated order-type ω, and the postulated cardinal number $\aleph_0$, are really entitled to rank as ordinal and cardinal numbers respectively. When we consider the ordinal number Ω, and the cardinal number $\aleph_1$, the state of the case is very different. In order that the existence of Ω might be on a parity with that of ω, it would require to be shewn that it is possible to define a set of objects, say points of the linear continuum, which should be such that, to each prescribed ordinal number of the second class, there corresponds a definite point of the continuum, *i.e.* to shew that a norm is possible which would define a set of points of order-type Ω. This has hitherto not been accomplished, nor have aggregates having any of the cardinal numbers $\aleph_2, \aleph_3, \ldots$ been defined by means of sets of rules. If it be urged that the postulation of the order-type Ω, and of the corresponding cardinal number $\aleph_1$, does not of itself lead to contradiction, it may be replied that such postulation does not entitle Ω and $\aleph_1$ to rank as numbers, in the sense in which

ω and $\aleph_0$ are numbers; for, in the latter case, the essential elements in the original conceptions of ordinal and cardinal numbers are all present, whereas this has not been shewn to be true of Ω and $\aleph_1$. Moreover, the postulation of the existence of Ω and $\aleph_1$, if it does not of itself lead to contradiction, can only be made by means of a principle which, when applied systematically, certainly leads to contradiction. In accordance with the criterion laid down above, $\aleph_1, \aleph_2, \ldots$ cannot, at the present time, be regarded as definite entities, and could not be regarded as in any true sense numbers, even if any meaning could be assigned to them.

It may conceivably turn out, in the future, to be possible to justify the postulation of the existence of certain of the numbers $\aleph_1, \aleph_2, \ldots$, together with the classes of ordinal numbers which would belong to them. It will, however, certainly never be possible to do so for the whole class $\{\aleph_\beta\}$, where β is any ordinal number of the aggregate of all ordinal numbers, in accordance with Cantor's complete scheme, because such postulation leads to unavoidable contradiction. The setting up of a scale of standards, to some of which standards no aggregates not consisting of the preceding numbers conform, involving, as it does, the employment of sphinx-like aggregates, to each of which no other aggregates can be shewn to be similar, would *à priori* appear to be an illegitimate extension of the notion of number, an essential element having dropped out; and *à posteriori* it has been shewn to lead to contradiction.

It may be urged that no contradiction would ensue if, in single instances, the existence of order-types and powers, considered to be definite entities, were postulated for aggregates of the unique character referred to above. But if this were done, such order-types and powers would not be entitled to rank as numbers; and such sporadic creations would be of no importance in Mathematical theory. Systematic postulation of this character is just what has been shewn to lead to a self-contradictory scheme of entities, and is therefore illegitimate.

A cardinal number has been defined* by B. Russell, to be a class of equivalent aggregates; it may then be urged that such class may contain only one member, and that this is sufficient for the existence of the cardinal number. In fact, Russell infers† the existence of the number $n+1$, from that of the numbers $0, 1, 2, 3, \ldots n$.

* *Principles of Mathematics*, vol. I, pp. 111—116.

† *Ibid.* p. 497. Since Russell regards the activities of the mind as irrelevant in questions of existence of entities, his view, and that here advocated, have no premisses in common. An advantage claimed for the conception of the nature of number, here advocated, over that of Russell, is that it does not lead to such a contradiction as that pointed out by Burali-Forti. Russell objects (see p. 114) to the conception of a number as the common characteristic of a family of equivalent aggregates, on the ground that there is no reason to think that such a single entity exists, with which the aggregates have a special relation, but that there may be many such

In accordance with the view here advocated of the nature of number, this definition, or any other one which allows the existence of a cardinal number to be inferred solely from the existence of a unique aggregate, to which no other aggregates have been shewn to be equivalent, must be rejected.

158. The conclusions at which we have arrived in the course of the above discussion, may now be summarized as follows:—

(1) The aggregates

$$1,\ 2,\ 3, \dots n, \dots \omega,\ \omega+1, \dots \Omega, \dots \beta, \dots$$

$$\aleph_0, \aleph_1, \aleph_2, \dots \aleph_n, \dots \aleph_\omega, \aleph_{\omega+1}, \dots \aleph_\Omega, \dots \aleph_\beta, \dots$$

of all ordinal numbers, and of all aleph-numbers, in the sense in which Cantor contemplates them, have no existence. Their existence cannot be established without the assumption of the principle that every normally ordered aggregate necessarily has a definite order-type, and a definite cardinal number, which can themselves be regarded as objects capable of being elements of an aggregate. This principle leads to contradiction, and must therefore be rejected as not being a universally valid truth.

(2) Of the aleph-numbers, the postulation of the existence of $\aleph_0$ has hitherto alone been justified*, by shewing that it is possible to define aggregates consisting of objects other than the ordinal numbers themselves, of which it is the characteristic cardinal number. The numbers $\omega, \omega+1, \dots \omega.2, \dots \omega^2, \dots \omega^\omega, \dots$ of the second class exist, but it has not yet been shewn that the totality of all such numbers, taken in order, has a definite order-type or a definite cardinal number; even if it be legitimate to speak of these numbers as forming a totality. To do this it would be necessary to shew that a finite set of rules can be set up which will suffice to define a definite object corresponding to each ordinal number of the second class.

(3) The existence of individual aleph-numbers, other than $\aleph_0$, with the classes of ordinal numbers belonging to them, may, in the future, be established; but it is not possible that this should be done beyond some definite stage.

It thus appears that there is at present no sufficient reason for thinking that any unenumerable aggregate is capable of being normally ordered.

It may be observed that an aggregate which consists wholly of distinct physical objects which do not penetrate one another must be enumerable: for each such object occupies some definite volume in space; and it has been shewn that any set of distinct portions of space is enumerable. It follows that the objects contained in an unenumerable aggregate, must,

entities, and that there are in fact an infinite number of them. The mind does, however, in point of fact, in the case of finite aggregates at least, recognize the existence of such a single entity, viz. the number of the aggregates; and this is a valid creation of our mental activity, subject to the law of contradiction.

* In *Math. Annalen*, vol. LX, p. 183 in a paper "Ueber wohlgeordnete Mengen," Schönflies has expressed a view which is to a certain extent in agreement with that here stated.

with the possible exception of an enumerable component of the aggregate, consist of ideal or abstract objects.

159. The regarding of a collection as a "whole" has been emphatically declared by Cantor, to be essential to the notion of an aggregate. It is no doubt true that, in a certain sense, every logical class, or aggregate as defined in § 153, forms a whole, as being dominated by a certain norm; but for the purposes of Mathematical Science, the fundamental question is, under what circumstances such an aggregate may be regarded as having a definite cardinal number, and if ordered, a definite order-type. This question has been fully discussed in the case of normally ordered aggregates; and the condition for an affirmative answer in the case of any other aggregate is of a similar character, viz. that it be possible to define other aggregates which have either the relation of similarity or that of equivalence with the given one.

Ordered aggregates have been defined, which are not normally ordered; and of such aggregates, the most important is the arithmetic continuum, defined in § 128, as of order-type θ. The justification for regarding θ as a definite object, with a definite cardinal number, must, as has been pointed out in § 158, be regarded as due to a postulation, subject to the law of contradiction. It has been seen that a class of aggregates exists which are similar to the linear continuum, and thus conform to the type θ, and have c as their common cardinal number; and this is in accordance with the regulative principle which we have maintained to be essential to justify our regarding c as a number.

As has been already remarked, aggregates may be defined, which are unordered. In such cases no question arises as to the existence of an order-type; but there is no reason why such aggregates should not have cardinal numbers, provided that in the case of such an aggregate equivalent aggregates can be found, the cardinal number in question being then their common characteristic. The aggregates of which the cardinal number is $f \equiv c^c$, are an example of this species of aggregate.

Two aggregates which have been independently defined are not necessarily comparable with one another, as regards either order-type or cardinal number. It cannot be assumed *à priori*, that the cardinal number of one of them is necessarily either greater, equal to, or less than that of the other, in the sense in which these relations have been defined in § 112. Further, it cannot be assumed, that an ordered aggregate, such as, for example, the continuum, is necessarily capable of being normally ordered. Two aggregates of abstract objects, which have been independently defined, may belong, no doubt, to the same universe of thought; but nevertheless, any particular category of relations may be too narrow to formulate any nexus between the two systems; so that it is conceivable that, so far as such relations as those of order, or cardinal number, are concerned, the two aggregates may be completely isolated from one another.

160. In some proofs of theorems which have been given by writers on this subject, which proofs have for their object the establishment of relations of inequality or equality of cardinal numbers, aggregates are employed, the elements of which are regarded as being successively defined by an infinite number of separate acts of choice. When we leave the region of the finite it would however appear that we have passed beyond the region in which definitions by arbitrary acts of choice can be regarded as adequate specifications of definite objects; and the existence of a norm would appear to be essential to our right to regard an aggregate as really defined, and therefore to justify our making use of the conception of such an aggregate in the proof of a theorem. The point may be illustrated by a discussion given* by Du Bois Reymond, in which he contemplates the existence of a number represented by a non-terminating decimal, in which the figures are determined by no law. He contemplates each figure in the decimal as being fixed by a throw of dice, and rejects the conception of such a decimal, (ewig gesetzloses Decimal), as representing a real number. A non-finite, or endless, process can be conceived of as a completed whole, only when it is subject to some kind of norm; thus a non-terminating decimal represents a number, only under the presupposition that a set of rules can be given, which would suffice to determine the figure that occupies any assigned place in the decimal. In general, the proof of the possibility of giving a norm is required before an aggregate of any particular character can be contemplated as existing, or can be legitimately made use of in a demonstration.

Cantor, in his proof (see § 117), that $\aleph_0$ is less than any other cardinal number, has assumed that it is possible to pick out of any given transfinite aggregate an enumerable component. This proof can only be accepted as valid in case it is possible to define an enumerable component of the aggregate in question. In a large class of cases, perhaps in all which are of importance in Mathematics, this condition can be satisfied; for example in the case of the continuum. In the aggregate of "all propositions," for example, the enumerable component might be taken to be that aggregate of propositions which asserts the existence of the numbers 1, 2, 3, ...

G. H. Hardy has extended† Cantor's method, for the purpose of shewing that every cardinal number is either an aleph-number, or is greater than all the aleph-numbers, and in particular that $2^{\aleph_0} = c \geqq \aleph_1$. This proof runs as follows:—Having given any aggregate whose cardinal number is $> \aleph_0$, we can choose from it successive individuals $u_1, u_2, \ldots u_\omega, \ldots u_\beta, \ldots$, corresponding to all the numbers of the first and second classes; and if the process came to an

* *Allgemeine Functionentheorie*, p. 91.

† *Quarterly Journal of Math.*, vol. xxxv, 1903, p. 88.

end, the cardinal number would be $\aleph_0$. Its cardinal number is therefore $\geqq \aleph_1$; and if $> \aleph_1$, $\geqq \aleph_2$, and so on. And if $> \aleph_n$, for all finite values of n, it must be $\geqq \aleph_\omega$; for we can choose individuals from the aggregate corresponding to all the numbers of the first, second, ... nth, ... classes. And by a repetition of these two arguments, we can shew that, if there is no $\aleph_\beta$ equal to the cardinal number of the aggregate, it must be at least equal to the cardinal number of the aggregate of all $\aleph_\beta$'s, and so greater than any $\aleph_\beta$.

Apart altogether from the question as to what constitutes all the aleph-numbers, this argument could only be valid, if it were shewn how the successive individuals $u_1, u_2, \ldots u_\beta, \ldots$ are to be defined by means of some norm, and also how the individuals of the aggregate which may correspond to the numbers of the first, second, ... nth, ... classes can be assigned by a norm. The process can neither come to an end, nor be regarded as, in any sense, a completed one, unless this has been done.

In connection with the definition, given in § 116, of the aggregate obtained by covering one aggregate by another one it must be assumed that each particular element of the aggregate of coverings is defined by a norm. This point will be exemplified in the discussion, which will be given in Chapter IV, of the cardinal number of all functions of a real variable.

In the theorem of § 120, the aggregate M_{AB} must be regarded as such that each element of it is defined by a norm. It is further necessary that the aggregate M'_{AB} be defined by a norm. This point may be illustrated by referring to the second proof in § 56, that $c > a$, which is a special case of the theorem of § 120. It is there hypothetically assumed that it is possible to define a number of the continuum corresponding to each integral number, by means of a norm; and thus the existence is assumed of a finite set of rules by means of which the nth figure of the number which corresponds to the integer n can be calculated. By introducing an additional rule, that, when this figure has been calculated, it is to be increased by unity, unless it be 9, in which case it is to be replaced by zero, the existence of a norm has been established, by which a number is defined that cannot correspond to any integer; and thus a contradiction is shewn to arise from the hypothesis made. It would not be sufficient to say that we may write down a number which differs, in at least one figure, from any of the numbers in the correspondence; it is essential to the validity of the proof, that such a number be shewn to be definable by a finite set of rules.

161. Two proofs have been advanced, that every cardinal number is necessarily an aleph-number; but this is equivalent to the statement that every aggregate which has a cardinal number can be normally ordered. If these proofs could be accepted as valid, the particular theorem would be established that the arithmetic continuum is capable of being normally ordered; and the only question which would remain open, as regards this aggregate, would be as to which particular aleph-number is the cardinal number of the continuum.

The first of these proofs, that of Jourdain*, is founded on the assumption that, if a cardinal number is greater than every aleph, there must be a part of the aggregate to which this cardinal number belongs, which can be made to have a (1, 1) correspondence with the "inconsistent" aggregate of all the ordinal numbers arranged in normal order. This assumption is regarded as justified by the process of making the successive elements of the aggregate of ordinal numbers correspond to elements of the given aggregate: it is then argued, that, if this process comes to an end, the cardinal number of the aggregate is an aleph; and that, if it does not come to an end, the given aggregate must contain a part that corresponds to the "inconsistent" aggregate of all the ordinal numbers; and thus that, in the latter case the aggregate is inconsistent, and has no cardinal number. The objection to this proof is the fundamental one which has been already stated, viz. that no norm is forthcoming by which the correspondence in question is defined; and, in default of such norm, there is no meaning in speaking of an essentially endless process as a completed one, or as having come to an end.

In the second proof, due† to E. Zermelo, no account is taken of the possibility that an aggregate may have no cardinal number, nor of the existence of "inconsistent" aggregates. The proof, which is fundamentally of a similar character to that of Jourdain, is represented as demonstrating that every aggregate can be normally ordered, and thus has an aleph as its cardinal number.

It is assumed that, in each part M' of a given aggregate M, one element m', called the special (ausgezeichnetes) element of M', can be chosen. A part M' must contain one element of M at least, and may contain all the elements; and the aggregate $\{M'\}$ of all parts of M is considered. Each element M' of $\{M'\}$ corresponds to a special element m' which belongs to M; and this particular mode of covering the elements of $\{M'\}$ by elements of M is called a "covering" γ; the employment of a particular "covering" γ, is essential to the proof. A γ-aggregate is then defined as follows:—Let M_γ be a normally ordered aggregate consisting of different elements of M, such that, if a be any arbitrarily chosen element of M_γ, and if A be the segment of M_γ defined by a, which segment consists of all the elements of M_γ that precede a, then a is always the special element of $M-A$. Every such aggregate M_γ is a γ-aggregate. If every element of M which occurs in a γ-aggregate be called a γ-element of M, it is shewn that the aggregate L_γ of all γ-elements can be so ordered that it is itself a γ-aggregate, and contains

* *Phil. Mag.* January 1904, pp. 67, 70.

† *Math. Annalen*, vol. LIX, 1904, "Beweiss, dass jede Menge wohlgeordnet werden kann."

all the elements of the original aggregate M. It follows then that M can be normally ordered.

Zermelo himself expressly recognizes the assumption made as to the existence of a definite "covering" γ. The objection to this assumption is of the same character as before, viz. that for its validity a norm must be shewn to be possible; this norm must assign to each part of the given aggregate a definite "special" element belonging to that part. In the case of such an aggregate as the continuum it is not clear how such a norm could be devised; indeed, it seems probable that a proof of the possibility of establishing such a norm involves difficulties comparable with those which occur in any attempt to prove the original theorem*. The non-recognition of the existence of "inconsistent" aggregates, which existence, on the assumption of Cantor's theory, cannot be denied, introduces an additional element of doubt as regards this proof. The aggregate L_γ, here employed, is parallel with the normally ordered aggregate which occurs in Jourdain's earlier proof.

162. As regards the method of G. H. Hardy (see § 151) for constructing a set of points of cardinal number $\aleph_1$, it was pointed out by Hardy, that an infinite freedom of choice arises in the case of each limiting number γ, since there are an indefinite number of sequences of the preceding ordinal numbers, of each of which sequences γ is the limiting number. Thus, for example, ω^2 is not only the limit of ω, $\omega.2$, $\omega.3, \ldots$, but also of $\omega+1$, $\omega.2+2$, $\omega.3+3, \ldots$. In the case of the smaller limiting numbers of the second class Hardy has shewn how to exercise this freedom of choice so as to obtain distinct sequences; thus ω is taken as the limit of $1, 2, 3, \ldots$; ω^2 is taken as the limit of ω, $\omega.2$, $\omega.3, \ldots$. In order however that the method should really suffice to define sequences of integers which shall correspond uniquely to each prescribed number of the first or of the second class, it would be necessary to replace this freedom of choice by a definite norm, or finite set of rules, which would decide, in the case of any particular limiting number γ, of what particular sequence of the preceding ordinal numbers γ must be regarded as the limit, for the purpose of forming the sequence of integers which is to correspond to it, in accordance with the mode of formation employed in the method.

Hardy has given no norm of this character, but has confined himself to the selection of the sequences which are to correspond to some of the lower limiting numbers of the second class. When we reach the region of the

* A criticism of Zermelo's proof has also been published by Borel, *Math. Ann.*, vol. LX, p. 194, and is substantially identical with the above, which was published in the *Proc. Lond. Math. Soc.*, ser. 2, vol. III. Borel however objects to the definition of an aggregate by an infinite number of acts of choice only when the aggregate is unenumerable; whereas the objection is really valid in the case of any non-finite aggregate.

ϵ-numbers of the second class, it is difficult, if not impossible, to imagine the nature of the norm which would suffice to make the decision referred to above; and no such norm is in fact forthcoming. On this ground, the method cannot be regarded as really defining a set of points such that a determinate point corresponds to each ordinal number of the first, or the second class.

163. In case the criticisms which have been given above, of the general theory of classes of order-types and of aleph-numbers, be accepted as wholly, or in part, valid, nevertheless the debt which Mathematical Science owes to the genius of G. Cantor will be in no material respect diminished. The fundamental distinction between enumerable and unenumerable aggregates, the interpretation of the arithmetic doctrine of limits, the ordinal theory of the arithmetic continuum, and the conception of the transfinite ordinal numbers of the second class, with their application to the theory of sets of points, remain as permanent acquisitions which rest upon a firm logical basis. This order of ideas has already become indispensable, for purposes of exact formulation, in Analysis and in Geometry; it is constantly receiving new applications, owing to its admirable power of providing the language requisite for expressing results in the theory of functions with the highest degree of rigour and generality. Cantor's creations have rendered inestimable service in formulating the limitations to which many results in Analysis, formerly supposed to be universally valid, are really subject. The outlying parts of the theory, to which exception has been taken, would not appear to be comparable in importance, for the general purposes of Analysis, with those parts to which the criticisms made are not applicable. The latter involve only a natural extension of the notion of Number, in which account is taken of all the elements that are essential to the conception of number in its original form; whereas we have endeavoured to shew that the more speculative general theory of aleph-numbers, and order-types, depends upon an extension of the notion of number which leaves out of account an essential element of that conception, viz. the notion of correspondence; and that this is the origin of the contradiction which arises when an endeavour is made to contemplate the totality of these new entities. The criticisms contained in the latter part of the present Chapter are advanced with some diffidence, on account of the great logical difficulties of the subject, and especially on account of the philosophical difficulties relating to existential propositions. It is hoped, however, that they may, in any case, be of utility as a contribution towards the discussion of questions of great interest which, at the present time, cannot be regarded as having been decisively settled.

The fact that the general theory of the aleph-numbers has received no applications in the theory of functions, and has indeed remained a purely abstract development of the theory of order, differentiates it from the theory of normally ordered enumerable aggregates, which has now become an essential instrument in the theory of functions of one or more variables. All aggregates of points in a continuum, which we at present know how to define, have either the power of the aggregate of rational numbers, or else that of the arithmetic continuum itself. The theories of these two kinds of aggregates, including, as they do, a complete arithmetic theory of limits, would thus appear to afford a sufficient basis for the development of Analysis.

CHAPTER IV.

FUNCTIONS OF A REAL VARIABLE.

164. If we suppose that an aggregate of real numbers is defined, the aggregate being either enumerable or of the power of the continuum, such an aggregate is said to be the domain of a real variable. It is necessary for the purposes of Analysis to be able to make statements applicable to each and every real number of the aggregate, and which shall be valid for any particular number that may at will be selected. This is done by employing the *real variable*, denoted by some symbol other than those used to denote real numbers; and the essential nature of the variable consisting in its being identifiable with any particular number of its domain. The symbols used for denoting variables differ from those employed in the case of numbers in being non-systematic. Operations involving real variables $x, y, z, \ldots$, with or without particular numbers, are carried out in conformity with the same formal laws as hold in the arithmetic of real numbers. The result of any such operation is itself a variable with a domain of its own, which may or may not be identical with that of any one of the constituent variables.

The numbers being used to designate in the usual manner the points of a set on a straight line, the variable may then be taken to refer to the points of the set.

If the given set of points be bounded, in the sense explained in § 46, then the domain of the variable is said to be *limited*. When the domain of the variable is not limited, it is said to be *unlimited* in one or in both directions.

The variable is said to be continuous in a given interval (a, b) when all the points of the interval, including a and b, belong to the domain of the variable. If the points a, b do not belong to the domain, but every internal point of the interval does so belong, the variable is said to be continuous in the *open* interval (a, b), or *within* the interval (a, b). It is unnecessary to give in detail the corresponding definitions applying to the case of an aggregate of any number n of dimensions, which is regarded as the domain of n independent variables $x_1, x_2, \ldots x_n$.

The term "variable" has been commonly associated with the conception of a point moving in a straight line or in a curve. It has however been pointed out in the course of the discussions of the continuum, contained in the earlier chapters, that the continuum cannot legitimately be regarded as a synthetic construction formed by a set of points determined successively. Successive determination is applicable only in the case of any enumerable sequence which may be defined within the continuum, and such a sequence may represent a succession of positions of a point moving in a straight line. It is however unnecessary to proceed to a detailed analysis of the conception of motion, because the Theory of Functions has no need of the conception of temporal succession. The theory makes continual use of simply infinite sequences determined in the continuum; and any such sequence may be regarded as a series of distinct determinations of the variable in which the elements are in logical succession, each element after the first being preceded and succeeded by definite elements.

THE FUNCTIONAL RELATION.

165. If to each point of the domain of the independent variable x there be made in any manner to correspond a definite number, so that all such numbers form a new aggregate which can be regarded as the domain of a new variable y, this variable y is said to be a (single-valued) *function* of x. The variables x, y are called the independent and the dependent variable respectively; and the functional relation between these variables may be denoted symbolically by the equation $y = f(x)$. In this definition no restriction is made *à priori* as regards the mode in which, corresponding to each value of x, the value of y is assigned; and the conception of the functional relation contains nothing more than the notion of determinate correspondence in its abstract form, free from any implication as to the mode of specification of such correspondence. In any particular case, however, the special functional relation must be assigned by means of a set of prescribed rules or specifications, which may be of any kind that shall suffice for the determination of the value of y corresponding to each value of x. Such rules may in any particular case be embodied in a single arithmetic formula from which the value of y corresponding to each value of x is arithmetically determinable; or the rules may be expressed by a set of arithmetic formulae each one of which applies to a part of the domain of the independent variable. In case these formulae be reducible to a set of mutually independent formulae, that set must be a finite one. In case the function be defined by an enumerably infinite set of formulae, each applicable to a part of the domain, these formulae cannot be mutually independent, but must be subject to some norm.

It must be observed that, when for any particular value of x the corresponding value of y is given by means of an arithmetic formula, the numerical value of y is in general only formally determinate; for in practice only a finite number of elements of a convergent aggregate which defines the value of y can in general be actually found, and thus the value of y can be specified only to any required degree of approximation, but it is still regarded as perfectly determinate.

The domain of x consisting of a set (P) of points, the values of y, in the case of a given functional relation $y=f(x)$, may be represented by points Q on another straight line, all such points forming a set (Q). The set (Q) is said to be the functional image of the set (P), determined by the function $f(x)$; to each point of (P) there corresponds a single point of (Q), if $f(x)$ be a single-valued function, but to each point of (Q) there may correspond a finite or an infinite number of points of (P).

The perfectly general definition of a function which has been given above is the culmination of a process of evolution which has proceeded largely in connection with the study of the representation of functions by means of trigonometrical series. By the older mathematicians a function was understood to mean a single formula, at first usually only a power of the variable; but afterwards it was regarded as defined by any one analytical expression, and was extended by Euler to include the case in which the function is given implicitly by a formal relation between the two variables. In connection with the problem of the determination of the forms of vibrating strings, which led to the discussion of functions represented by trigonometrical series, the conception arose of a single function defined in different intervals by means of different analytical expressions. The arbitrary nature of a function given by a graph was distinctly recognised by Fourier; thus the notion of a function was emancipated from the restriction that an *à priori* representation of it by a single formula is necessary.

The idea that a function can be defined completely, in the case when the domain of the independent variable is a finite continuous interval, by means of a graph arbitrarily drawn, leaves out of account the essentially unarithmetic nature of geometrical intuition. A curve that is drawn is indistinguishable by the perception from a sufficiently great number of discrete points; and thus all that is really given by an arbitrarily drawn graph consists of more or less arithmetically inexact values of the ordinates at those points of the x-axis at which we are able to measure ordinates. In order that a curve may be really known, sufficiently to serve for the purpose of defining a function, a series of rules must be prescribed, by means of which the values of the ordinates can be formally determined at all points of the x-axis. It is sometimes said, in order to illustrate the generality of the functional relation, that a function is definable in the form of a table which specifies values of y

corresponding to values of x, this table being of a perfectly arbitrary character. The inadequacy of such illustration is manifest, if we consider that even if the table were an endless one, as has been remarked in § 160, no aggregate of y-values can be defined by an endless set of numbers, apart from the production of a norm by which those numbers are defined. Moreover, even if the table were subject to a definite norm, it could only theoretically suffice to define a function of a variable whose domain consisted of an enumerable set of points, and would be totally inapplicable to the case in which the variable has a continuous domain, unless some special restrictive assumptions as to the nature of the function be introduced, by means of which the values of the function are made determinate at the remaining points of the continuous domain.

It thus appears that an adequate definition of a function for a continuous interval (a, b) must take the form first given to it by Dirichlet*, viz. that *y is a single-valued function of the variable x, in the continuous interval (a, b), when a definite value of y corresponds to each value of x such that $a \leqq x \leqq b$, no matter in what form this correspondence is specified.* A particular function is actually defined when y is arithmetically defined for each value of x.

No elaborate theory is required for functions which retain their complete generality, in accordance with the abstract definition given above, since no deductions of importance can be made from that definition which will be valid for all functions. When, however, the nature of a function is in some way restricted, either in the whole domain, or in the neighbourhoods of special points of that domain, there is room for the development of a theory which shall deal with the peculiarities that follow from such restrictions upon the complete generality of functions.

166. The functions defined in accordance with the above definition are known as *single-valued* functions, since, to each value of x in the domain of x, there corresponds a single value of y. The definition may be so generalised as to be applicable to *multiple-valued* functions. This is done by replacing the requirement that, to each value of x in the domain of x there shall correspond a single value of y, by the more general statement that, to each value of x there shall correspond a definite aggregate of values of y. The aggregate of values of y may, for any particular value of x, consist of a finite, or of an infinite, set of numbers. A particular function is then defined when the aggregate of values of y is arithmetically determinate for each value of x, in accordance with the criteria for the determinacy of a linear aggregate which have been developed in the theory of aggregates. Although the Theory of Functions, as developed in the present work, is mainly concerned with single-valued functions, it is necessary, or at least convenient, in the

* See Dirichlet's *Werke*, vol. I, p. 135.

course of the examination of particular functions and classes of functions, to make use of auxiliary functions which are multiple-valued at certain points of the domain of the independent variable. Moreover, Dirichlet's definition, in its original form, has the inconvenience that it excludes from the category of functions such as are represented by analytical expressions which, for particular values of the independent variable, cease to define a single number. For example, an infinite series which, for particular values of the variable, either diverges, or ceases to converge to a single definite limit, does not define a single-valued function in accordance with Dirichlet's definition, for the whole domain of the variable, and yet it is convenient to so extend the meaning of the term function that a function may be nevertheless defined for the whole domain by such a series.

The distinction has been considered in detail by Brodén* between those functions for which the relation between the dependent variable y and the independent variable x is formally the same for the whole domain of x, and those functions for which the domain of x is divisible into a plurality of parts, for which the forms of the relation between x and y are different. He remarks that the distinction is one relating to the character of the definitions rather than to the nature of the functions themselves; in the former case the function is said to be *homonomically* defined, and, in the latter case, to be *heteronomically* defined. Brodén has given a formal proof that, when a function is heteronomically defined, the number of parts into which the domain of x is divided, so that the relations of y to x in any one part are completely independent of the relations in the other parts, must be finite.

The Theory of Functions of a Real Variable is concerned with the classification of functions, according as they possess various special properties, *e.g.* continuity, differentiability, integrability, throughout the domain of the independent variable, or at, or near, special points which form part of that domain. The theory requires the introduction of precise arithmetical definitions of the scope and meaning of these characteristic properties, and is largely concerned with the determination of criteria which shall suffice to decide, in the case of a function defined in some special manner, what can be inferred as regards the possession by such function of properties other than those that are immediately apparent from the definition itself. Much of the theory is concerned with a minute examination of functions, and of classes of functions, which possess properties that do not occur in the case of those functions which are employed in ordinary analysis and in its applications to Geometry and Physics; and the theory has in consequence frequently been described as the Pathology of Functions. It appears however from the theory itself that many of those peculiarities, which from the point of view of traditional Analysis would be described as ex-

* *Acta Univ. Lund.* vol. XXXIII, 1897, "Functionentheoretische Bemerkungen und Sätze."

ceptional, have no claim to be so described; that in fact it is in the functions of ordinary Analysis that the abnormalities really occur, such functions occupying an exceptional position in relation to a scientific Analysis of the properties of functions in general. An important result of the labours of those who have developed the modern theory of functions of a real variable has been that restrictive assumptions, which had previously been unconsciously made in the processes of ordinary Analysis, have been placed in a clear light; and it has been shewn that modes of reasoning which had their origin in an uncritical application of ideas obtained from intuition would fail to yield correct results when applied to cases of sufficient generality, the unsoundness of the logical basis of such reasoning being thereby demonstrated.

In ordinary Analysis the domain of the independent variable is taken to be a limited, or unlimited, continuous interval. In the theory of functions, on the other hand, it has been found advantageous to consider also the properties of functions defined for a domain which is not a continuous one. It appears, in particular, that a non-dense perfect set of points, or more generally any closed set, is well suited to be the domain of a function, inasmuch as, for such domains, the principal peculiarities of functions, such as continuity, differentiability, &c., are capable of precise formulation, and can serve for purposes of classification, exactly as in the case of functions defined for a continuous domain. Much of the recent progress in the subject is due to a recognition of the parity of all perfect sets of points, not only as regards their internal structure, but also in relation to their fitness for forming the domains for which functions can be defined, without loss of any of the characteristic properties that serve for the classification of functions of a real variable, or of several such variables.

EXAMPLES.

1. A function $f(x)$ may be defined for the interval $(0, 1)$ as follows:—for $\frac{1}{n} \geqq x \geqq \frac{1}{n+1}$, $f(x) = \frac{1}{n}x^2$, and for $x=0$, $f(0)=1$, n denoting any positive integer. In this case, the norm by which the function is defined is expressible by an enumerable set of formulae which are however not independent of one another.

2. A function may be defined as follows:—for $1 \geqq x \geqq \frac{1}{2}$, $f(x) = \frac{x}{2}$; for $\frac{1}{2} \geqq x \geqq \frac{1}{3}$, $f(x) = \frac{x}{3}$; for $\frac{1}{3} \geqq x \geqq \frac{1}{4}$, $f(x) = \frac{x}{5}$, ..., and in general, for $\frac{1}{n} \geqq x \geqq \frac{1}{n+1}$, $f(x) = \frac{x}{P_n}$, where P_n denotes the nth of the prime numbers 2, 3, 5, 7, ... If the function is to be defined at the point $x=0$, this may be done by assigning to $f(0)$ any arbitrarily chosen value we please. It will be observed that the values $\frac{x}{P_n}$ are in this case not representable by a single expression which involves n and x only.

3. Any number x of the interval (0, 1) except 0, can be uniquely expressed in the form

$$\frac{b_1}{2}+\frac{b_2}{2^2}+\ldots+\frac{b_n}{2^n}+\ldots,$$

where b_n has for every value of n one of the values 0, 1, and it is stipulated that all the b_n are not to be zero from and after any fixed value of n.

A multiple-valued function* may be defined by $y=x^{\frac{1}{n}}$, where n has all positive integral values for which $b_n=1$. This is a homonomic definition, although no analytical expression of a unitary character can be given for the representation of y.

THE UPPER AND LOWER LIMITS OF FUNCTIONS.

167. A function $y=f(x)$, being defined for the domain of x, we have seen that the values of y form a set of points, determined as usual upon a straight line, which is called the functional image of that set of points which forms the domain of x. In case the set of points, which represent the values of y, is a bounded set, the function $f(x)$ is said to be *limited* in the domain of x.

When the set of values of y is bounded, either boundary may be a limiting point, or only an extreme point, of the set. For convenience, and in accordance with usage, the terms upper limit and lower limit will be applied to denote the upper and lower boundaries of y, even when the boundary is not a limiting point of the set, but is only an extreme point, without being a limiting point in the sense in which this term is used in the theory of sets of points. Thus we may say that:—

If the set of points y, which represents the functional image of a function $f(x)$, defined for a given domain of x, have an upper and a lower boundary, then the function $f(x)$ is said to be a limited function, and the boundaries are said to be the upper and lower limits of $f(x)$ in the domain of x.

The upper or the lower limit of a function $f(x)$ in its domain may or may not be *attained*, *i.e.* there may or may not be a value of x, in the domain of x, for which the functional value is equal to the upper, or to the lower limit, of the function. An upper or lower limit, which is attained, is an extreme point of the set of values of y, and may or may not be a limiting point of such set, in the accurate sense. An upper or a lower limit which is not attained is certainly a limiting point of the set of values of y.

In case y have no upper limit, or no lower limit, for the domain of x, the function $f(x)$ is said to be *an unlimited function*. In this case there exist values of the function, of one or of both signs, which are numerically greater than any arbitrarily assigned number A.

When y has no upper limit in the domain of x, the function is said to have the improper limit $+\infty$, in the domain of x. Similarly, when y has no

* Brodén, *loc. cit.* p. 4.

lower limit, it is said to have the improper limit $-\infty$. It is frequently said, for the sake of brevity, that the upper or the lower limit of the function is infinite.

The excess of the upper limit of a function, in its domain, over its lower limit, is called the fluctuation (Schwankung) *of the function in the domain.*

In case the upper or the lower limit is infinite, the function is said to have an infinite fluctuation in its domain.

Instead of the whole of the domain of x, we may consider that part which lies in a given interval (a, b), including the end-points a and b, and the preceding definitions may be applied to this portion of the domain; thus:—

The upper limit of a function $f(x)$ in an interval (a, b) is the upper limit of the function when only those points of the domain of x which lie in (a, b) are taken into account. A similar definition applies to the case of the lower limit.

The excess of the upper limit of $f(x)$, in the interval (a, b), over its lower limit in that interval, is called the fluctuation of $f(x)$ in the interval (a, b).

In case one or both of the limits is infinite, the fluctuation of the function in (a, b) is said to be infinite.

If the upper limit of $f(x)$ in (a, b) is attained, *i.e.* if there exists a value c of x such that $f(c)$ is the upper limit, where c is a point of the domain in (a, b), then this upper limit is said to be the *upper extreme* of the function in (a, b); and a similar definition applies to the *lower extreme.*

If the end-points a, b of the interval be left out of account, in case they belong to the domain of x, the fluctuation is called the fluctuation in the open interval (a, b). This is sometimes spoken of as the *inner fluctuation* of the function in (a, b), and is determinable as the limit of the fluctuation in the interval $(a+\epsilon, b-\epsilon)$, when ϵ is indefinitely diminished.

168. In accordance with the definition which has been given for a function in any domain, the value of the function at any particular point of the domain has a definite finite value. It may happen that a point P, of the domain of x, may be such that in any arbitrarily small neighbourhood of P either the upper or the lower limit of the function, or both, may not exist; so that, however small the neighbourhood of P may be chosen, there exist functional values in that neighbourhood which are numerically greater than any number that may be assigned. In that case, the point P is said to be *an infinity, or point of infinite discontinuity of the function*; although the function has a definite finite value at the point P itself.

Although $f(x)$ is not properly defined at a point $P, (x_0)$, unless a definite numerical value be assigned to $f(x_0)$, nevertheless an improper definition of the functional value at the point P is sometimes admitted, of the form $\frac{1}{f(x_0)}=0$;

in this case the function is said to possess an *infinity* at P. This infinity is said to be *removable* provided that, when the functional value at P is altered to some finite value, the function have finite upper and lower limits in a sufficiently small neighbourhood of P.

There are other cases in which an improper definition of the functional value at a point x_0 of the domain of x is admitted. The function may be defined by means of an infinite series, of which the terms are given functions of x. This series may diverge at the particular point x_0; but it is nevertheless frequently convenient to regard the series as defining the function for all values of x in some interval which includes x_0. The functional value at x_0 is then regarded as infinite.

In accordance with strict arithmetic theory, the function is regarded as undefined at points where no definite finite value of the function is specified. For the most part, in the theory which will be here developed, this restriction will be rigidly adhered to. It will be found, however, that in cases, such as in the theory of infinite series, in which it is convenient to admit improper definitions of functions at particular points, no essential change in the main results of the theory will have to be made.

In some cases it will be found convenient to remove the restriction that at each point of the domain of the independent variable the function shall be single-valued, and to define the function in such a manner that, at single points, or at each point of some set belonging to the domain of x, the function may possess finite or infinite multiplicity. It will be found, in the cases in which it is convenient to make this extension of the meaning of a function, that no difficulty arises as regards the use of results primarily applicable to functions which are single-valued at all points of the domain of the variable, without exception.

THE CONTINUITY OF FUNCTIONS.

169. Let the domain of the independent variable x be continuous, and either bounded or unbounded; and denote the function y at the point x by $f(x)$.

The function $f(x)$ is said to be continuous at the point a of the domain of x, if, corresponding to any arbitrarily chosen positive number ϵ whatever, a positive number δ dependent on ϵ can be found, such that $|f(a+\eta)-f(a)|<\epsilon$, for all positive or negative values of η which are numerically less than δ, and which are such that $a+\eta$ is in the domain of x. At an end-point of a limited domain, the values of η will have one sign only.

In accordance with this definition, a neighbourhood $(a-\delta, a+\delta)$ of the point a exists, such that the function, at any point in the interior of this interval, differs numerically from its value at a, by less than ϵ. It follows that the inner fluctuation of the function in $(a-\delta, a+\delta)$ is less than 2ϵ, and it is obvious that the fluctuation in any interval interior to $(a-\delta, a+\delta)$ is less than 2ϵ. The condition of continuity of the function $f(x)$ at the point a may thus be stated to be that *a neighbourhood of the point can be found in which the fluctuation of the function is as small as we please.*

The above definition of continuity at a point is that due to Cauchy, and is a particular case of the definition of continuity for a function of any number of variables. If we denote by $f(x, y, z, \ldots)$ a function of the variables $x, y, z, \ldots$ defined for any continuous domain, the condition of continuity at the point $(\alpha, \beta, \gamma, \ldots)$ is that, corresponding to every arbitrarily chosen positive number ϵ, a number δ dependent on ϵ, can be found, such that $|f(\alpha+h, \beta+k, \gamma+l, \ldots)-f(\alpha, \beta, \gamma, \ldots)|<\epsilon$, provided $h, k, l, \ldots$ have any values which are numerically less than δ. In this case, a neighbourhood $(a-\delta, a+\delta)$ of a point of a linear domain, is replaced by a "rectangular cell," which is a square in the case of a two-dimensional domain. The definition of continuity has been stated by Heine* in a form which depends upon the notion of a convergent sequence of numbers or of points. Let $(P_1, P_2, \ldots P_n, \ldots)$ be a convergent sequence of points in the given domain, and of which P is the limiting point. The condition of continuity of the function at P is that, for every such convergent aggregate which has P as limiting point, the numbers $f(P_1), f(P_2), \ldots f(P_n), \ldots$ form a convergent sequence which represents the number $f(P)$. That this definition is equivalent to Cauchy's is seen at once by considering a sequence of values of ϵ which have the limit zero, and are such that $\epsilon_1 > \epsilon_2 > \epsilon_3 \ldots$.

A function which is not continuous at a point a may satisfy the condition that in a neighbourhood of a on the right the fluctuation of the function may be made as small as we please by taking the neighbourhood small enough; the function is then said to be *continuous on the right at a*. A similar definition applies to continuity on the left.

A function is said to be *continuous in the interval* (a, b) if it satisfies the condition of continuity at every point in the interval.

The function is said to be *in general* continuous in the interval, if, when arbitrarily small neighbourhoods of a finite number of points are removed, the function† be continuous in each of the remaining intervals. Either of the points a, b may be one of this finite number of points.

170. The domain of the independent variable has hitherto been considered to be continuous; it is however clear from a consideration of the definition of continuity, either in Cauchy's or in Heine's form, that the

* *Crelle's Journal*, vol. LXXIV, p. 182.

† C. Neumann uses the term abtheilungsweise stetig: see his work "Die nach Kreis, Kugel, und Cylinder-functionen fortschreitenden Reihen."

definition is applicable in case the domain of the independent variable is not continuous, but consists of any set of points which contains limiting points that belong to the set. It is, of course, only at such a limiting point that the question of continuity arises; for at an isolated point of the aggregate there are only a finite number of values of the function in any sufficiently small neighbourhood of the point. If P be a point of the domain of x which is a limiting point of the domain, the function is continuous at P when, for every sub-set $(P_1, P_2, \ldots P_n, \ldots)$, all the points of which belong to the domain, and which has P as limiting point, the numbers $f(P_1), f(P_2), \ldots f(P_n), \ldots$ form a convergent sequence of which $f(P)$ is the limit. If the function be continuous at every limiting point of the domain of x it is said to be continuous relatively to the given domain; and thus the notion of continuity of a function is applicable whatever be the domain of the independent variable, except when it consists of an isolated set of points.

Let $P_1, P_2, P_3, \ldots$ be a convergent sequence of points of the domain of x, of which P_ω is the limiting point; and let P_ω also belong to the domain of x. Supposing the functional image, corresponding to $f(x)$, to contain the points $Q_1, Q_2, Q_3, \ldots$ which correspond to $P_1, P_2, P_3, \ldots$, let $Q_1, Q_2, Q_3, \ldots$ form a convergent sequence of which the limiting point Q_ω corresponds to P_ω. If this condition be satisfied, however the convergent sequence be chosen in the domain of x, the aggregate (Q), of values of y, is said to be a continuous functional image of the domain (P) of x.

It is clear that the continuous functional image of a closed domain is itself closed. For, corresponding to the points of a convergent sequence $(Q_1, Q_2, Q_3, \ldots)$, in (Q), there corresponds an aggregate $(P_1, P_2, P_3, \ldots)$, in (P), which must have at least one limiting point, and all such limiting points belong to the domain (P), and must correspond to the limiting point of $(Q_1, Q_2, Q_3, \ldots)$, which therefore belongs to the aggregate (Q). Moreover if (P) be perfect, the continuous functional image (Q) is perfect also; for, corresponding to any particular point Q' of (Q), we may take a point P' of (P), for which Q' is the image. P' is the limiting point of a convergent sequence of points of (P), and to this convergent sequence there corresponds a convergent sequence in (Q), of which Q' is the limiting point. It has thus been shewn that (Q) contains no isolated points, and therefore (Q) is perfect.

If (Q) be a continuous functional image of the closed set (P), and if only one point of (P) correspond to each one point of (Q), then (P) is a continuous functional image of (Q).

To the points of any convergent sequence $(Q_1, Q_2, \ldots)$ in (Q), of which Q_ω is the limiting point, there corresponds a convergent sequence $(P_1, P_2, P_3, \ldots)$ in (P) of which P_ω is the limiting point, and P_ω is the functional image of Q_ω.

171. The theorem has been given by Weierstrass that, *if* (a, b) *be any interval containing points of the domain of a function, then one point at least exists in the interval, which is such that, in any arbitrarily small neighbourhood of that point, the upper limit of the function is the same as the upper limit of the function in the whole interval* (a, b).

This theorem holds for all functions without restriction, and it makes no difference whether the whole interval (a, b), or only a set of points in that interval, belongs to the domain of the independent variable.

If M denotes the upper limit of the function in (a, b), the case of an indefinitely great upper limit being included, let the interval be divided into a number n of equal parts. It is then clear that the upper limit of the function for no one of these parts can be greater than M, and that, in one at least (α_1, β_1) of these sub-intervals, the upper limit of the function must be M. Divide (α_1, β_1) into n equal parts, then, as before, one of these parts (α_2, β_2), at least, is such that M is the upper limit of the function in it. Proceeding in this manner, we obtain a sequence $(\alpha_1, \beta_1), (\alpha_2, \beta_2), \dots (\alpha_r, \beta_r), \dots$ of intervals whose lengths converge to zero, such that each one is contained in the preceding one, and such that M is the upper limit of the function in any one of these intervals. In accordance with the theorem of § 48, there is one point x_1, which is in all these intervals; and this point x_1 is such that in any arbitrarily small neighbourhood the upper limit of the function is M. A similar result holds for the lower limit of a function.

In the case of a function which is continuous in the interval (a, b), it follows from the foregoing theorem that the upper and lower limits of the function in (a, b) are both finite, and thus that *a function which is continuous in an interval is limited in that interval.*

For consider that point x_1 in (a, b), in the arbitrarily small neighbourhood of which $(x_1 - \epsilon, x_1 + \epsilon)$ the upper limit has the same value as for the whole interval (a, b). Since the function is continuous at x_1, corresponding to a given number δ, a number ϵ can be found such that $|f(x) - f(x_1)| < \delta$, provided x lies in $(x_1 - \epsilon, x_1 + \epsilon)$; consequently the upper limit of $f(x)$ in this interval must be finite, and hence $f(x)$ has a finite upper limit in (a, b). It may be shewn in a similar manner that the function has a finite lower limit.

A function which is continuous in the interval (a, b) *is such that its upper limit and its lower limit are each actually attained at one point at least in the interval, i.e. the function has an upper extreme and a lower extreme in the interval.*

For suppose, if possible, that $f(x_1)$ has a value A different from M; and consider an arbitrarily small interval $(x_1 - \epsilon, x_1 + \epsilon)$ for which M is the upper limit of the values of the function; then points can be found in this interval for which the function differs by less than an arbitrarily small number δ from M. These values of the function would differ from $f(x_1)$ by an amount which is not arbitrarily small, and this would be inconsistent with the condition of continuity of the function at the point x_1. It follows that we must have $f(x_1) = M$. Similarly it may be shewn that the lower limit m is reached at least once in the interval (a, b).

CONTINUOUS FUNCTIONS DEFINED FOR A CONTINUOUS INTERVAL.

172. It will now be shewn that *if $f(x)$ be continuous in the continuous domain (a, b), and if $f(a)$, $f(b)$ have opposite signs, then there is at least one value of x in the interval, for which $f(x)$ vanishes.*

Suppose $f(a) < 0$, $f(b) > 0$; then on account of the continuity of the function we know that at a point x, for which $f(x)$ is negative, an interval $(x, x + \epsilon)$ can be found for which the fluctuation of the function is as small as we please; and therefore the interval can be so chosen that for every point of it the function is negative. Dividing the whole interval into any n equal parts consider the signs of the function at the points of division. Writing e for $1/n$, let $a + (p_1 + 1)e(b - a)$ be the first of these for which the function is positive; thus for $a + p_1 e(b - a)$ the function is negative, or zero. Divide the interval $\{a + p_1 e(b - a),\ a + (p_1 + 1)e(b - a)\}$ into n equal parts, and suppose the point of division, $a + (p_1 e + p_2 e^2)(b - a)$, is the last of these, reckoned to the right, for which the function is negative, or zero. Proceeding in this manner we obtain a series of numbers S_m, where S_m denotes $p_1 e + p_2 e^2 + \ldots + p_m e^m$, which are such that for $a + S_m(b - a)$ the function is negative, or zero; and for $a + (S_m + e^m)(b - a)$ the function is positive. Let c be the limit of the sequence $a + S_m(b - a)$; then it can be shewn that $f(c) = 0$. For if $f(c)$ were negative, then an interval $(c, c + \delta)$ could be found, for all points of which the function is negative: and by choosing m sufficiently great, the point $a + (S_m + e^m)(b - a)$ could be made to fall within the interval $(c, c + \delta)$, for which point the function would be positive: hence $f(c)$ cannot be negative. Again, $f(c)$ cannot be positive; for in that case an interval $(c - \delta, c)$ can be found for all points of which the function is positive; but by choosing m large enough the point $a + S_m(b - a)$ can be made to fall within this interval, and then for this point the function is negative, or zero. Since then the function $f(c)$ cannot be either positive or negative it must therefore be zero.

From this theorem we can deduce that, *whatever values $f(a)$, $f(b)$ may have, there must be in the interval (a, b) at least one value of x, for which $f(x)$ has any prescribed value lying between $f(a)$ and $f(b)$.*

Let this value be C, and suppose $f(a) < C < f(b)$; then the function $f(x) - C$ is continuous in the given interval, is negative when $x = a$, and positive when $x = b$; thus it vanishes at least once in the interval (a, b).

A continuous function has frequently been defined as a function such that, if $f(a), f(b)$ be its values at any two points a and b, then the function passes through every value intermediate between $f(a)$ and $f(b)$, as x changes from a to b. The property contained in this definition has been shewn above to hold of every function which is continuous in accordance with Cauchy's definition; but the converse theorem does not, in general, hold. The definition just referred to is accordingly not equivalent to that of Cauchy, which is here adopted as the basis of the treatment of continuous functions. As an example of the non-equivalence of the two definitions, we may consider the function defined by $y = \sin\frac{1}{x}$, for $x \gtrless 0$, and by $y = 0$, for $x = 0$. For this function there are values of x between a and b for which $f(x)$ has any assigned values c lying between $f(a)$ and $f(b)$; but the function is not continuous, in accordance with Cauchy's definition, in any interval (a, b) which contains the point 0. It is, in fact, easily seen that the point 0 is a point of discontinuity of the function; for an arbitrarily small neighbourhood of the point 0 contains points at which the function has all values in the interval $(-1, 1)$.

As another example* of a function which satisfies the condition referred to, but is discontinuous in accordance with Cauchy's definition, let the number x in the interval $(0, 1)$ be expressed as a decimal $\cdot a_1 a_2 a_3 \ldots a_n \ldots$; then consider the decimal $\cdot a_1 a_3 a_5 a_7 \ldots$. If this last decimal is not periodic, we take $f(x) = 0$; if it is periodic, and the first period commences at a_{2n-1}, we take $f(x) = \cdot a_{2n} a_{2n+2} a_{2n+4} \ldots$. The function so defined for the interval $(0, 1)$ of x has every value between 0 and 1, in every arbitrarily small interval in the domain of x; thus the function is discontinuous at every point. A value of x for which $f(x)$ has any prescribed value $\cdot p_1 p_2 \ldots p_n \ldots$ is

$$\cdot a_1 A a_3 B \ldots K a_{2n-1} p_1 a_{2n} p_2 a_{2n+2} \ldots,$$

where $\cdot a_1 a_3 \ldots$ is any periodic decimal, the first period of which begins at a_{2n-1}, and $A, B, \ldots K$ are arbitrarily chosen digits. Nevertheless there are values of x between α and β at which the function takes any assigned value intermediate between $f(\alpha)$ and $f(\beta)$.

CONTINUOUS FUNCTIONS DEFINED AT POINTS OF A SET.

173. It will now be shewn that, *if a function $f(x)$, having prescribed values at each point of an infinite set of points in the interval (a, b), be continuous in that interval, then the values of the function are determinate at each point of the derivative of the set.*

Suppose $\alpha_1, \alpha_2, \alpha_3 \ldots \alpha_n \ldots$ to be a convergent sequence of points, for which x_1 is the limiting point, and suppose $f(\alpha_1), f(\alpha_2) \ldots f(\alpha_n) \ldots$ to be known; it will be shewn that these functional values form a convergent sequence whose limit is $f(x_1)$. An interval $(x_1 - \delta, x_1 + \delta)$ can always be found, corresponding

* See Lebesgue, *Leçons sur l'intégration*, p. 90.

to any fixed number ϵ, such that the function at any point of this interval differs from $f(x_1)$ by less than the arbitrarily small number ϵ; this follows from the continuity of the function. A number n can be found such that all the points $\alpha_n, \alpha_{n+1}, \alpha_{n+2}, \ldots$ lie within the interval $(x_1 - \delta, x_1 + \delta)$. It follows that $|f(x_1) - f(\alpha_n)|$ and $|f(x_1) - f(\alpha_{n+1})| \ldots$, etc. are all less than ϵ, which is arbitrarily small; hence $f(x_1)$ is the limit of the sequence $f(\alpha_1), f(\alpha_2) \ldots$, and thus $f(x_1)$ is determinate. From this special case it follows that, for all the limiting points of a given set of points in (a, b), the values of the continuous function are determinate. It further appears that the function is determinate for all points which belong to any derivative of the given set, for the points of which set the values of the function are known.

In particular, *if a continuous function have prescribed values for points of a set which is everywhere-dense throughout the interval* (a, b), *then its values are determinate for all points of the interval.*

A special case of such a set would be all the rational points within the interval. It follows that a continuous function whose values are known for all the rational points in an interval is determinate for all the irrational points. A continuous function which is known to be constant for all the rational points has the same constant value for all the irrational points in the interval.

A generalization of the above theorem is, that a function which is continuous with reference to a domain which consists of a set (P), and is known for all points of a sub-set which is everywhere-dense in (P), is determinate for every point of (P). This may be seen by considering that every point of (P) is a limiting point of the sub-set, and applying the same reasoning as before.

174. From the theorem established above, that a continuous function is determinate when its values at an everywhere-dense enumerable set of points are prescribed, we may deduce that *the cardinal number of the aggregate of all continuous functions of a real variable is the cardinal number* c *of the continuum.*

We may suppose the values of a function to be prescribed at the rational points. The cardinal number of the aggregate of all functions defined for the rational points only is the cardinal number of the ways of covering the aggregate of rational numbers by the aggregate of numbers of the continuum. This number is c^a, which has been shewn in § 148 to be equal to c. Only some of the "coverings" of this kind are such as will give rise to continuous functions; hence the aggregate of all continuous functions is a part of the aggregate of all possible coverings of the set of rational numbers by the numbers of the continuum. It follows that the cardinal number of the aggregate of all continuous functions is $\leqq c$. Again, this cardinal number is $\geqq c$; for among the continuous functions are those each of which is constant and everywhere equal to any assigned number of the continuum; and thus the aggregate of all continuous functions contains a part which has the cardinal number c. Since the cardinal number is $\leqq c$, and also $\geqq c$, it is equal to c.

It has been shewn by Borel* that the aggregate of all continuous functions and also that of all analytical functions of two or more variables have the cardinal number c.

The cardinal number of the aggregate of all functions of a real variable is that of all coverings of the continuum by itself; this is, in accordance with the definition in § 116, denoted by c^c, for which we may write f.

Each particular "covering" of the numbers of the continuum by themselves is definable by a definite norm, and corresponding to each such covering there is a definite function for a continuous domain. Let the aggregate of all such functions be denoted by F: it will then be proved that the cardinal number f, of F, is $> c$.

First, F has a part which is equivalent to the continuum. This is at once seen, since the functions $f(x) = c$, where c is any number of the continuum, constitute such a part. It follows that $f \geqq c$.

Next, let it be assumed, if possible, that F is equivalent to a part of the continuum. As has been just proved, such a part cannot have a cardinal number $< c$; we therefore assume that F is equivalent to the set of numbers of the continuum. This amounts to the assumption that F can be ordered in the same type as the continuum, so that, to any assigned number ξ of the continuum, there corresponds a definite set of rules R_ξ which defines a function $f_\xi(x)$. The correspondence between ξ and R_ξ must itself be defined by a set of rules, so that when ξ is assigned, R_ξ, and therefore the function $f_\xi(x)$ is defined. The aggregate $\{f_\xi(x)\}$ must contain every definable function of a real variable. The number ξ being assigned, $f_\xi(x)$ is producible, and its existence implies that, at any assigned point ξ', the functional value $f_\xi(\xi')$ can be determined arithmetically. We may take, for example, $\xi' = \xi$; and thus, if ξ is assigned, $f_\xi(\xi)$ is known. We may regard $f_\xi(\xi)$ as a function of ξ; for its value at any point ξ can be arithmetically determined, and it is therefore an element of the aggregate F of all functions. With this understanding as to $f_\xi(\xi)$, choose a fixed number, say unity, then the function $\phi(\xi) \equiv f_\xi(\xi) + 1$ has a definite norm; for we have only to add to the rules by which $f_\xi(\xi)$ is defined, the further rule, that, at each point ξ, unity is to be added to the value of $f_\xi(\xi)$. We have now a new definable function $\phi(x)$; but this cannot possibly belong to the aggregate F, for if it do so belong, there must be some one point ξ_1 of the continuum, with which it corresponds; but $\phi(x)$ cannot be identical with $f_{\xi_1}(x)$, for $\phi(\xi_1)$ and $f_{\xi_1}(\xi_1)$ differ by unity. Since $\phi(\xi)$ is not contained in F, contrary to the hypothesis, it follows that F cannot be equivalent to the continuum, and thus the theorem, $f > c$, is established. It has therefore been shewn that $f > c$, and consequently that—*the aggregate of all functions of a real variable has a cardinal number f, greater than c.*

* See *Leçons sur la théorie des fonctions*, p. 127.

UNIFORM CONTINUITY.

175. It will now be shewn that if the domain of x be a continuum, then a continuous function is ***uniformly continuous*** through the domain of x; that is to say, a number $\bar{\delta}$ can be found corresponding to any given ϵ, such that, for all values of x, the fluctuation of $f(x)$ within the neighbourhood $(x-\bar{\delta}, x+\bar{\delta})$, or for all of this neighbourhood which lies within the domain, is less than the number ϵ.

If within $(x-\delta, x+\delta)$ the fluctuation of the function be less than ϵ, then for a given ϵ, the number δ must have an upper limit $\phi(x, \epsilon)$, which is in general a function of x, and is essentially positive. It must be shewn that $\phi(x, \epsilon)$, for the whole domain of x, has a finite lower limit which is not zero; this lower limit is then a suitable value for $\bar{\delta}$. If the lower limit of $\phi(x, \epsilon)$ be zero, there must be at least one point of the domain, such that for its arbitrarily small neighbourhood, the lower limit of $\phi(x, \epsilon)$ is zero. Suppose, if possible, x_1 to be such a point; then the values of $\phi(x, \epsilon)$, for a convergent sequence of points whose limiting point is x_1, must form a convergent sequence with zero for lower limit. Since $f(x)$ is continuous at x_1, a neighbourhood $(x_1-\delta', x_1+\delta')$ can be found, with δ' finite, such that the fluctuation of the function within that neighbourhood is less than ϵ; it follows that, for any point x of the interval $(x_1-\frac{1}{2}\delta', x_1+\frac{1}{2}\delta')$, a neighbourhood $(x-\frac{1}{2}\delta', x+\frac{1}{2}\delta')$ exists within which the fluctuation of the function is less than ϵ. This is contrary to the hypothesis that $\phi(x, \epsilon)$ becomes arbitrarily small by taking x near enough to x_1; thus the lower limit of $\phi(x, \epsilon)$ cannot be less than $\frac{1}{2}\delta'$, and is therefore finite.

It has thus been shewn* that it is unnecessary to draw a distinction, as has sometimes been done, between functions which are uniformly, and those which are non-uniformly, continuous in the continuous domain of x; for all continuous functions are uniformly continuous.

The theorem may also be stated in the following form:—

If $f(x)$ be continuous in the interval (a, b), then, corresponding to any arbitrarily chosen positive number ϵ, a number η can be determined, such that the condition $|f(x_1)-f(x_2)|<\epsilon$, is satisfied, where x_1, x_2 are any two points in (a, b), such that $|x_1-x_2|<\eta$.

The following theorem can be immediately deduced:—

If a function be continuous in a finite interval, then the interval can be divided into a finite number of sub-intervals in every one of which the fluctuation of the function is less than a prescribed positive number.

* This theorem was first stated and proved by Heine; see *Crelle's Journal*, vol. LXXI (1870), p. 361, and vol. LXXIV (1872), p. 188.

It is in fact clear that, if ϵ be the prescribed number, the condition is satisfied when the interval is subdivided in any manner such that the length of the greatest of the sub-intervals is $< \eta$.

Another proof of the above theorem, in an extended form, will be given in § 185, by employing the Heine-Borel theorem.

It is clear that the above proof applies also to the case in which the domain of x is not a continuum, but is a perfect aggregate, or any closed aggregate; because the essential point of the proof depends upon the limiting points all belonging to the aggregate. For aggregates which are not closed the proof does not apply; thus a function which is continuous, relatively to an aggregate which is not closed, is not necessarily uniformly continuous.

THE LIMITS OF A FUNCTION AT A POINT.

176. Let a be a limiting point of the set of points which forms the domain of the independent variable x; the point a may or may not itself belong to the domain of x. Let $(a, a+h)$ be a neighbourhood of a on the right, and let $U(h)$, $L(h)$ denote the upper and lower limits of a given function $f(x)$ for all the points of the domain of x which are interior to the interval $(a, a+h)$. It will be observed that $f(a)$, if it exists, is not reckoned amongst the functional values of which $U(h)$, $L(h)$ are the upper and lower limits.

Let a descending sequence of values be assigned to h, which converges to zero; denoting this sequence by $h_1, h_2, h_3, \ldots$, the corresponding numbers $U(h_1), U(h_2) \ldots U(h_n) \ldots$ form a sequence of which the members do not increase, and therefore they have in general a definite lower limit, which is called the *upper limit of $f(x)$ at a on the right*. It may happen that all the upper limits $U(h)$ are infinite, in which case we say that the upper limit of $f(x)$ at a on the right is $+\infty$; or it may happen that the sequence $U(h_1), U(h_2) \ldots U(h_n) \ldots$ has no lower limit, in which case we say that the upper limit of $f(x)$ at a on the right is $-\infty$. In any case, the finite or infinite upper limit of $f(x)$ at a on the right is denoted by $\overline{f(a+0)}$.

The numbers $L(h_1), L(h_2) \ldots L(h_n) \ldots$ form a sequence of which the elements do not diminish, and they have in general a definite upper limit, which is called the *lower limit of $f(x)$ at a on the right*, and may, as in the former case, have infinite values ∞ or $-\infty$. This limit is denoted by $\underline{f(a+0)}$.

Corresponding definitions apply to the left of the point a; and the limits of $f(x)$ at a on the left are denoted by $\overline{f(a-0)}$, $\underline{f(a-0)}$ respectively. In case the point a is a limiting point of the domain of x on one side only, the two limits of the function at a on the other side are non-existent.

The definitions may be stated shortly as follows:—

The upper limit $\overline{f(\alpha+0)}$ *of a function at* α *on the right is the limit of the upper limit of* $f(x)$ *in the open interval* $(\alpha, \alpha+h)$, *as* h *is indefinitely diminished.*

The lower limit $\underline{f(\alpha+0)}$ *of a function at* α *on the right is the limit of the lower limit of* $f(x)$ *in the open interval* $(\alpha, \alpha+h)$ *when* h *is indefinitely diminished.*

The definitions for the left of α may be stated in a precisely similar manner.

It is to be observed that the four functional limits $\overline{f(\alpha+0)}$, $\underline{f(\alpha+0)}$, $\overline{f(\alpha-0)}$, $\underline{f(\alpha-0)}$ are entirely independent of $f(\alpha)$, in case α belongs to the domain for which $f(x)$ is defined. Any arbitrary alteration in the value of $f(\alpha)$ will not affect these four limits of $f(x)$ at α.

The conditions that the point α may be a point of continuity of the function $f(x)$ are that $\overline{f(\alpha+0)}$, $\underline{f(\alpha+0)}$, $\overline{f(\alpha-0)}$, $\underline{f(\alpha-0)}$, $f(\alpha)$ must all have the same finite value.

It may happen that $f(\alpha)$, $\overline{f(\alpha+0)}$, $\underline{f(\alpha+0)}$ have one and the same finite value, but that either or both of $\overline{f(\alpha-0)}$, $\underline{f(\alpha-0)}$ may not have this value; in that case $f(x)$ is said to be continuous at α on the right. Continuity at α on the left is defined in a similar manner.

If the four functional limits at α be all finite and equal, but $f(\alpha)$ have a different value, then the function is said to have a *removable discontinuity* at the point α. In this case the function would be made continuous at α merely by properly altering the value of $f(\alpha)$.

The four functional limits at the point $x=0$ are usually denoted by $\overline{f(+0)}$, $\underline{f(+0)}$, $\overline{f(-0)}$, $\underline{f(-0)}$ respectively.

177. If the upper and lower limits of $f(x)$ at α on the right have the same value, this common value is called *the limit of* $f(x)$ *at* α *on the right*, and is denoted* by $f(\alpha+0)$. If the upper and lower limits of $f(x)$ at α on the left have the same value, this is called the limit of $f(x)$ at α on the left, and is denoted by $f(\alpha-0)$. Both of the limits on the right or left at a point, when such limit exists, may be either finite or infinite.

The limit at $x=0$, on the right, is denoted by $f(+0)$; and the corresponding limit on the left is denoted by $f(-0)$.

The limit at a point P on one side may be also defined as follows:—Let $(P_1, P_2, P_3 \ldots)$ be any convergent sequence of points belonging to the domain of x, which is such that α, or P, is its limiting point, and such that all the

* This notation was introduced by Dirichlet; see *Werke*, vol. I, p. 156.

points of the sequence are on the one side of P. The values of $f(x)$ at $P_1, P_2, P_3 \dots$ form an aggregate which may be a convergent sequence; let us suppose it to be so, and also that its limit has a value which is independent of the particular sequence, which is however subject to the conditions above stated. In that case this limit is denoted by $f(\alpha+0)$, or by $f(\alpha-0)$, as the case may be, and is called the limit of $f(x)$ at α, on the right or left.

It may be observed that the necessary and sufficient condition for the existence of a definite finite limit on the right at α is that, corresponding to every arbitrarily small number ϵ, a neighbourhood $(\alpha, \alpha+\delta)$ can be found, such that the difference of the values of the function at every pair of points of the domain of x, which are in the interior of this interval, is numerically less than ϵ.

The necessary and sufficient condition that $f(\alpha+0)$ should exist and $=+\infty$, is that, if A be an arbitrarily chosen positive number, then δ can be so determined that at every point interior to $(\alpha, \alpha+\delta)$, the condition $f(x)>A$ is satisfied. In order that $f(\alpha+0)$ may exist and $=-\infty$, the corresponding condition is that $f(x)<-A$.

It is possible that one of the limits $f(\alpha+0)$, $f(\alpha-0)$ may exist and not the other. If the domain of x be either a continuum or a perfect set, α may be taken to be at any point of the domain.

When the condition for the existence of $f(\alpha+0)$ or of $f(\alpha-0)$ at a point α is not satisfied, the convergent sequence $(P_1, P_2, \dots P_n \dots)$, of which $P(\alpha)$ is the limiting point, may be such that $f(P_1), f(P_2), \dots f(P_n) \dots$ is either not a convergent sequence, or else that its limit depends upon the particular choice of the points $P_1, P_2, \dots P_n \dots$. In this case the fluctuation of $f(x)$ within an arbitrarily small neighbourhood $(\alpha, \alpha+\delta)$ on the one side of α is either a finite number which has not zero for its limit when δ is indefinitely diminished, or else it is indefinitely great, however small δ may be.

178. If $x_1, x_2, x_3, \dots x_n \dots$ be a convergent sequence of points belonging to the domain of x, with α for its limiting point, then the sequence $f(x_1), f(x_2), \dots f(x_n) \dots$ may not be convergent; but, if it be convergent, its limit may have (1) a single value independent of the mode in which the convergent sequence is chosen, in which case α is either a point of continuity of $f(x)$, or a point of removable discontinuity of the function; or (2) one of two values, in which case both the limits $f(\alpha+0)$, $f(\alpha-0)$ exist; or (3) one of a finite, or an indefinitely great, number of values, which all lie between the greatest and least of the four functional limits at α.

The aggregate of all possible values of the limits of the convergent sequences $f(x_1), f(x_2), \dots f(x_n) \dots$, corresponding to different convergent sequences $(x_1, x_2, \dots x_n \dots)$, with α as limiting point, is called ***the aggregate*** *of functional limits* (Werthevorrath) at the point a. The aggregate of functional limits will be shewn, in § 190, to be necessarily a closed set.

179. If the domain of x be unbounded in one or in both directions, it may happen that a point x_1 of the domain can be found, corresponding to every arbitrarily chosen positive number ϵ, such that the difference between the values of $f(x)$, for any two values of x which are both greater than x_1, is numerically less than ϵ. In this case the function has a definite limit as x is increased indefinitely in its domain; and this is called the limit of $f(x)$ for $x = \infty$. Under a corresponding condition $f(x)$ may have a definite limit for $x = -\infty$.

If, as x increases, a point x_1 of the domain of x, corresponding to each assigned positive number A chosen as great as we please, can be found, such that $f(x) > A$ for all values of x which belong to the domain and are $> x_1$, then the limit of $f(x)$ is said to be ∞, as x is increased indefinitely. If $f(x) < -A$, for all such values of x_1, then the limit of $f(x)$ is said to be $-\infty$. Similar definitions apply to the case in which x has indefinitely great values in the negative direction.

In case the limit $f(\alpha + 0)$, at a point α on the right, do not exist as a definite number, and be not infinite with a fixed sign, it is frequently convenient to regard $f(\alpha + 0)$ as still existent, but indeterminate, and capable of all values belonging to some closed set of which $\overline{f(\alpha + 0)}$, $\underline{f(\alpha + 0)}$ are the extreme values. It is then said that $f(\alpha + 0)$ is indefinite in value, and that $\overline{f(\alpha+0)}$, $\underline{f(\alpha + 0)}$ are its *limits of indeterminacy*. A similar remark applies to $f(\alpha - 0)$, which may also be either definite, or indefinite, with $\overline{f(\alpha - 0)}$, $\underline{f(\alpha - 0)}$ as its limits of indeterminacy. One or both of the limits of indeterminacy, in either case, may be infinite.

THE DISCONTINUITIES OF FUNCTIONS.

180. Let us suppose the domain of x to include all points in a sufficiently small neighbourhood of a point a; or, in any case, let a be a limiting point of the domain of x.

The fluctuation of the function $f(x)$ in the neighbourhood $(\alpha - \delta, \alpha + \delta)$ of the point α depends in general upon δ, but cannot increase as δ is diminished. It therefore has a lower limit for values of δ which converge to zero. This limit, which may be zero, finite, or indefinitely great, is called the saltus (Sprung), or measure of discontinuity, of the function $f(x)$ at α; thus:—

The saltus, or measure of discontinuity, of a function $f(x)$ at a point α, is the limit of the fluctuation of the function in a neighbourhood $(\alpha - \delta, \alpha + \delta)$, as δ converges to zero.

The upper limit of the function $f(x)$ in the interval $(\alpha-\delta, \alpha+\delta)$ has a lower limit, as δ is indefinitely diminished, which is called the maximum of the function $f(x)$ at α.

The lower limit of the function in the same interval, has an upper limit, as δ is indefinitely diminished, which is called the minimum of $f(x)$ at α. Either the maximum or the minimum at a point may be indefinitely great.

The saltus of $f(x)$ at α is easily seen to be the excess of the maximum at α over the minimum.

It is clear that the maximum of $f(x)$ at α is the greatest of the numbers $\overline{f(\alpha+0)}$, $\overline{f(\alpha-0)}$, $f(\alpha)$, and that the minimum is the least of the numbers $\underline{f(\alpha+0)}$, $\underline{f(\alpha-0)}$, $f(\alpha)$; and thus that the *saltus* at α is the excess of the greatest over the least of the numbers $\overline{f(\alpha+0)}$, $\underline{f(\alpha+0)}$, $\overline{f(\alpha-0)}$, $\underline{f(\alpha-0)}$, $f(\alpha)$.

At a point of continuity of $f(x)$, the saltus is zero. Any point at which the saltus has a finite value, or is indefinitely great, is called a *point of discontinuity* of $f(x)$, and in the latter case it is said to be a *point of infinite discontinuity.*

If the neighbourhood $(\alpha, \alpha+\delta)$ on the right of α be taken, the lower limit of the fluctuation in this neighbourhood when δ is indefinitely diminished is called *the saltus at α on the right.* This is equivalent to the excess of the greatest over the least of the three numbers $\overline{f(\alpha+0)}$, $\underline{f(\alpha+0)}$, $f(\alpha)$. A corresponding definition applies to the saltus at α on the left.

181. The points of discontinuity of a function may be classified as follows:—
(1) If both the limits $f(\alpha+0)$, $f(\alpha-0)$ exist and have definite values which differ from one another, the point α is said to be a point of discontinuity *of the first kind*, or a point of *ordinary* discontinuity.

The difference between the greatest and least of the three numbers $f(\alpha+0)$, $f(\alpha-0)$, $f(\alpha)$ is the saltus, or measure of discontinuity, of the function at α. If α be not a point of the domain of x, $|f(\alpha+0)-f(\alpha-0)$ measures the saltus at α; and if α be a point of the domain, and $f(\alpha)$ lies between $f(\alpha+0)$ and $f(\alpha-0)$, then the saltus is also measured by $|f(\alpha+0)-f(\alpha-0)|$.

When $f(\alpha)$ does not lie between $f(\alpha+0)$ and $f(\alpha-0)$, the function is said to have an *external saltus* at α.

In every case, the saltus on the right is measured by $|f(\alpha+0)-f(\alpha)|$, and that on the left by $|f(\alpha-0)-f(\alpha)|$.

Whether there be an external saltus at α or not, the number $|f(\alpha+0)-f(\alpha-0)|$ is said to measure the *oscillation* (Schwingung) at α. The oscillation at a point differs from the saltus in that the functional value $f(\alpha)$ at the point is in the former case disregarded.

If $f(\alpha)=f(\alpha-0)$, whilst $f(\alpha)\neq f(\alpha+0)$, the function is said to be *ordinarily discontinuous at α on the right*. If $f(\alpha)\neq f(\alpha-0)$, whilst $f(\alpha)=f(\alpha+0)$, the function is said to have an *ordinary discontinuity at α on the left*.

It may happen that $f(\alpha+0), f(\alpha-0)$ have equal values which differ from $f(\alpha)$. In that case the discontinuity at α is said to be removable; since by merely altering the functional value at the one point α, the function can be made continuous at the point.

(2) If neither of the limits $f(\alpha+0), f(\alpha-0)$ exists, the discontinuity at α is said to be *of the second kind*.

The *oscillation** at α is measured by the excess of the greater of the numbers $\overline{f(\alpha+0)}$, $\overline{f(\alpha-0)}$ over the lesser of the two numbers $\underline{f(\alpha+0)}$, $\underline{f(\alpha-0)}$, the value of $f(\alpha)$ being left out of account.

The differences $\overline{f(\alpha+0)}-\underline{f(\alpha+0)}$, $\overline{f(\alpha-0)}-\underline{f(\alpha-0)}$ may be spoken of as the oscillation at α on the right, and on the left, respectively.

By Dini†, a definition of the saltus is adopted which differs from the one which we have employed; he takes the greatest of the four differences $|\overline{\underline{f(\alpha\pm0)}}-f(\alpha)|$ as the measure of the saltus, the greater of the two differences $|\overline{\underline{f(\alpha+0)}}-f(\alpha)|$ being taken as the measure of the saltus on the right.

(3) It may happen that one of the two limits $f(\alpha+0), f(\alpha-0)$ exists as a definite number, whilst the other does not. In this case the point α may be said to be a *point of mixed discontinuity*.

If $f(\alpha)$ exist and be equal to that one of the two limits $f(\alpha+0), f(\alpha-0)$ which exists, then the function is continuous at α on one side, and has a discontinuity of the second kind on the other side.

(4) If one or more of the four limits $\overline{\underline{f(\alpha\pm0)}}$ be indefinitely great, the point α is one of infinite discontinuity.

Under infinite discontinuities is sometimes included the case in which $f(\alpha)$ is defined by $1/f(\alpha)=0$, or when $f(x)$ is defined as the limiting sum of a series which, for the value α, becomes divergent.

182. In an arbitrarily small neighbourhood $(\alpha, \alpha+h)$, on the right of a point α at which the limits $\overline{f(\alpha+0)}, \underline{f(\alpha+0)}$ have different values, there must be an infinite number of points at which $f(x)>\overline{f(\alpha+0)}-\epsilon$, where ϵ is an arbitrarily small fixed number.

* This definition of the "Schwingung" is given by Pasch in his *Einleitung in die Differential- und Integralrechnung*, p. 189.

† See *Grundlagen*, p. 55.

For if there were only a finite number of such points in $(\alpha, \alpha+h)$, h could be chosen so small that all such points would be excluded from the neighbourhood; thus, in a sufficiently small neighbourhood $(\alpha, \alpha+h)$, we should have at every internal point $f(x) \leqq \overline{f(\alpha+0)} - \epsilon$; and thus the upper limit at α on the right could not be $\overline{f(\alpha+0)}$. In a similar manner it can be shewn that, in the arbitrarily small neighbourhood $(\alpha, \alpha+h)$, there must be an infinite number of points at which $f(x) < \underline{f(\alpha+0)} + \epsilon$.

In this case we say that, in the arbitrarily small neighbourhood of α on the right, the function makes an infinite number of finite oscillations. In case of the infinity of one or of both of the limits $\overline{f(\alpha+0)}$ and $\underline{f(\alpha+0)}$, and in the latter case if they be of opposite signs, the function makes an infinite number of infinite oscillations in the arbitrarily small neighbourhood of α. A similar remark applies to the case in which $\overline{f(\alpha-0)}$, $\underline{f(\alpha-0)}$ have unequal values. It has thus been shewn that:—

A point of discontinuity of the second kind is one such that, in its arbitrarily small neighbourhood, the function makes an infinite number of finite or infinite oscillations.

In an arbitrarily small neighbourhood on either side of a point of discontinuity of the first kind, the function may make an infinite number of oscillations; but since the neighbourhood can be chosen so small that the fluctuation of the function in its interior is arbitrarily small, the oscillations, when they are infinite in number, are arbitrarily small sufficiently near the point.

EXAMPLES.

1. Let $f(x)=\sin x/x$, when $x \neq 0$, and $f(x)=A$, when $x=0$.

In this case $f(+0)=f(-0)=1$, $f(0)=A$; thus $f(x)$ has a removable discontinuity at $x=0$, unless $A=1$, in which case the function is continuous in any interval.

2. Let $f(x)=\dfrac{1}{x-a}$; we have then $f(a+0)=\infty$, $f(a-0)=-\infty$, and $f(a)$ is undefined.

3. Let $f(x)=(x-a)\sin\dfrac{1}{x-a}$; then $f(a+0)=0$, $f(a-0)=0$. This function is continuous at $x=a$, and makes an infinite number of oscillations in any neighbourhood of that point.

4. If $f(x)=\dfrac{1}{x-a}\operatorname{cosec}\dfrac{1}{x-a}$, then

$$\overline{f(a+0)}=\infty,\ \underline{f(a+0)}=-\infty,\ \overline{f(a-0)}=\infty,\ \underline{f(a-0)}=-\infty.$$

This function has an infinite discontinuity of the second kind at the point a.

5. If $f(x)=e^{\frac{1}{x}}$, we have $f(+0)=\infty$, $f(-0)=0$. If $f(x)=\dfrac{1}{1-e^{\frac{1}{x}}}$, then $f(+0)=0$, $f(-0)=1$.

6. If $f(x)=\sin x$, $\lim_{x=\infty} f(x)$ is indeterminate, the limits of indeterminacy being $+1$, -1. In the case $f(x)=x\sin x$, the corresponding limits of indeterminacy are $+\infty$, $-\infty$.

7. Let $y=E(x)$, where $E(x)$ denotes the integral part of x. This function is discontinuous when x has an integral value n; we then have $E(n-0)=n-1$, $E(n)=n$, $E(n+0)=n$.

8. Let (x) denote the positive or negative excess of x over the nearest integer; and when x exceeds an integer by $\frac{1}{2}$, let $(x)=0$. This function is continuous except for values $x=n+\frac{1}{2}$, where n is an integer. We have $(n+\frac{1}{2})=0$, $(n+\frac{1}{2}-0)=\frac{1}{2}$, $(n+\frac{1}{2}+0)=-\frac{1}{2}$.

SEMI-CONTINUOUS FUNCTIONS.

183. If $\phi(x)$ be a function defined for a continuous domain, and if, corresponding to every arbitrarily chosen positive number ϵ, a neighbourhood $(x-h,\ x+h)$ of a particular point (x) can be determined such that for every point x' in this neighbourhood the condition, $\phi(x')<\phi(x)+\epsilon$, be satisfied; then the point x is said to be a point* of *upper semi-continuity* of the function $\phi(x)$.

If a neighbourhood of the point x can be determined, for each ϵ, such that $\phi(x')>\phi(x)-\epsilon$, then the point x is said to be a point of *lower semi-continuity* of the function $\phi(x)$.

That a point x may be a point of continuity of the function $\phi(x)$, it is necessary that both the above conditions be satisfied.

If every point of the domain (a, b), for which the function $\phi(x)$ is defined, is a point of upper semi-continuity, then the function $\phi(x)$ is said to be an *upper semi-continuous function.*

A similar definition applies to a *lower semi-continuous function.*

It is clear that, if $\phi(x)$ be a lower semi-continuous function, then $-\phi(x)$ is an upper semi-continuous function. Thus the properties of the one class of functions may easily be extended to the other class.

If $f(x)$ be a function defined for the interval (a, b), and if $\phi(x)$, $\psi(x)$ denote the maximum and the minimum of $f(x)$ at the point x, then $\phi(x)$ is an upper semi-continuous function, and $\psi(x)$ is a lower semi-continuous function.

For a neighbourhood $(x-h,\ x+h)$ of any point x can be determined, such that the maximum (see § 180) of $f(x)$ for every point in this neighbourhood is less than $\phi(x)+\epsilon$, where ϵ is a prescribed positive number. At every point in $(x-h_1,\ x+h_1)$ where h_1 is chosen $<h$, the value of the function ϕ is less than $\phi(x)+\epsilon$. Since this holds for every value of ϵ, the function $\phi(x)$ is upper semi-continuous at x.

* See Baire's memoir "Sur les fonctions des variables réelles," *Annali di mat.* series III^a, vol. III, 1899.

It is clear that the function $\psi(x)$, where $\psi(x)$ denotes the minimum of $f(x)$ at the point x, is a lower semi-continuous function.

The saltus $\phi(x)-\psi(x)$ of the function $f(x)$, may be taken to be the value of a function $\omega(x)$ which is called the *saltus-function* of $f(x)$.

The saltus-function $\omega(x)$ of any function $f(x)$ is an upper semi-continuous function.

For $\phi(x)$, $-\psi(x)$ are both upper semi-continuous functions, and it is easily seen that the sum of two such functions belongs to the same class.

If $\phi(x)$ be any upper semi-continuous function, then the set of points, for which $\phi(x) \geqq \alpha$, is a closed set, where α is any fixed number.

For let $P_1, P_2, P_3, \dots P_n \dots$ be a sequence of points at each of which the condition $\phi(x) \geqq \alpha$, is satisfied, and let P be the limiting point of the sequence. Let us suppose that, if possible, $\phi(P) < \alpha$; then a neighbourhood of P can be determined, such that at every point in it the value of $\phi(x)$ is less than α, and hence this neighbourhood cannot contain any of the points of the sequence; but this is contrary to the hypothesis that P is the limiting point of the sequence. It follows that $\phi(P) \geqq \alpha$, and thus that the set of points, for which $\phi(x) \geqq \alpha$, contains all its limiting points, and is therefore a closed set.

It can be shewn, in a similar manner, or it can be deduced from the above theorem, that *the set of points for which $\psi(x) \leqq \alpha$, is a closed set; where $\psi(x)$ is any lower semi-continuous function.*

If we apply the theorem proved above to the saltus-function $\omega(x)$, of any function $f(x)$, we obtain the following theorem:—

Having given any function $f(x)$ defined for a continuous domain, the saltus-function $\omega(x)$ is such that the set of points for which $\omega(x) \geqq \alpha$, forms a closed set.

184. *If $\phi(x)$ be an upper semi-continuous function, and if at every point of the domain of x the minimum of $\phi(x)$ be zero, then there exists a set of points, everywhere-dense in the domain of x, at which $\phi(x)$ is itself zero.*

For, in any interval (α, β), the minimum of $\phi(x)$ is zero, and therefore a point P in the interior of (α, β) can be found at which $\phi(P) < \frac{1}{2}\epsilon$, where ϵ is a prescribed positive number. Since $\phi(x)$ is an upper semi-continuous function, an interval (α_1, β_1) interior to (α, β), and containing P in its interior, can be determined, such that for every point x in it, $\phi(x) < \phi(P) + \frac{1}{2}\epsilon < \epsilon$. Similarly, it can be shewn that an interval (α_2, β_2) interior to (α_1, β_1) can be found, such that at every point in (α_2, β_2) the condition $\phi(x) < \frac{1}{2}\epsilon$, is satisfied. Proceeding in this manner, we can determine a set of intervals $(\alpha_1, \beta_1), (\alpha_2, \beta_2) \dots (\alpha_n, \beta_n) \dots$, each of which is interior to the preceding one, and such that at every point x in (α_n, β_n), the condition $\phi(x) < \frac{1}{2^{n-1}}\epsilon$, is satisfied. This sequence of intervals, continued indefinitely, determines a point Q, in the interior of all of them, such that $\phi(Q) < \frac{1}{2^{n-1}}\epsilon$, for every value of n; and hence $\phi(Q) = 0$. It has thus been shewn that, in any interval whatever contained in the domain of the variable, there exists a point at which $\phi(x)$ is zero; and therefore the set of all such points is everywhere-dense in the domain of the variable.

In particular, we see that *if $\omega(x)$ be the saltus-function of any given function $f(x)$, and if $\omega(x)$ has its minimum equal to zero at every point of the domain of x, then $\omega(x)$ vanishes at an everywhere-dense set of points. The points of this set are the points of continuity of $f(x)$.*

185. The following theorem is a generalization* of the theorem of § 175, that a continuous function is uniformly continuous in its (closed) domain.

If a function $f(x)$ be defined for the interval (a, b), and if k be a number greater than the maximum of the saltus-function $\omega(x)$ in (a, b), then there exists a number α, such that within every interval in (a, b) of length not exceeding α, the fluctuation of $f(x)$ is $< k$.

The theorem of § 175 is the particular case which arises when the maximum of $\omega(x)$ is zero.

The theorem is most easily established by means of an application of the Heine-Borel theorem given in § 68.

For any point P in (a, b) a neighbourhood can be determined, such that the fluctuation of $f(x)$ within this neighbourhood is $< k$. If we conceive such a neighbourhood to be determined for each point in (a, b), then a finite number of these intervals can be chosen, such that every point in (a, b) is interior to one at least of the intervals. The end-points of this finite set of intervals form a finite set of points in (a, b); let then α be the smallest of the distances between consecutive points of this finite set. Any interval whatever in (a, b) of length not exceeding α is within one of the intervals of the finite set; hence within such an interval, the fluctuation of $f(x)$ is less than k.

The definitions given above, and the theorems established, are applicable when the domain of the variable x is not an interval (a, b), but any closed set of points. In that case, we regard functional values in any interval as only existing at those points in the interval which belong to the domain of x.

Moreover, the definitions, and theorems are applicable to the case of functions of a number n of variables. In this case, instead of an interval $(x - h, x + h)$ used in defining semi-continuous functions and the saltus-function, the "sphere"

$$(\xi_1 - x_1)^2 + (\xi_2 - x_2)^2 + \ldots + (\xi_n - x_n)^2 = h^2$$

* See Baire, *loc. cit.* p. 15.

may be employed. Corresponding to the interval $(x-h, x+h)$, we take the set of points $(\xi_1, \xi_2 \dots \xi_n)$, for which

$$(\xi_1 - x_1)^2 + (\xi_2 - x_2)^2 + \dots + (\xi_n - x_n)^2 \leq h^2.$$

EXAMPLE.

If* $f(x)$ be any function, $\phi(x)$ its maximum, and $\psi(x)$ its minimum at the point x, and $g(x)$ be any continuous function, then $\phi(x)+g(x)$, $\psi(x)+g(x)$ are the maxima and minima at x, of $f(x)+g(x)$. Also the functions $\phi(x)-f(x)$, $f(x)-\psi(x)$ have, each in any domain, the minimum zero.

THE CLASSIFICATION OF DISCONTINUOUS FUNCTIONS.

186. Let us suppose a function to be defined for all points in a continuous interval (a, b); at each point x the saltus of the function has a finite value, or is indefinitely great, its value being zero at a point of continuity. With a view to the classification of functions, in accordance with the distribution of the points of continuity and of discontinuity in the interval (a, b), the question arises, what is the most general distribution of the points of continuity?

The answer to this question is contained in the theorem:—

The points of continuity of a function, defined for a continuous interval, form an inner limiting set.

To prove this theorem, let ϵ be a fixed positive number, and enclose each point P of continuity of a function $f(x)$ in an interval so chosen that the fluctuation of $f(x)$ therein is less than ϵ; all the points of continuity are then enclosed in a set of intervals which in general overlap. Imagine these sets of intervals constructed corresponding to a sequence of diminishing values of ϵ which converges to zero; there exists then a set of points which are interior to intervals of all these sets of intervals, since this set of points includes all the points of continuity of $f(x)$. If Q be any point which belongs to the inner limiting set so defined, Q must be a point of continuity of $f(x)$; for corresponding to any arbitrarily small number ϵ_n, Q is in the interior of some interval in which the fluctuation of the function is less than ϵ_n, and thus Q is a point of continuity of the function.

In accordance with the theorems which have been obtained in § 97, relating to inner limiting sets, the points of continuity of a function may form an enumerable set which contains no component dense-in-itself, or else they form a set of the cardinal number of the continuum. In the latter case the set is of the second category, provided it be everywhere-dense.

* Baire, *loc. cit.* p. 9.

These results lead to the following classification* of functions:—

(1) A function may have no points of continuity, it is then said to be totally discontinuous.

(2) The points of continuity may form an enumerable set which has no component dense-in-itself.

(3) The set of points of continuity may be of the cardinal number of the continuum, and

(*a*) non-dense;

(*b*) everywhere-dense and unclosed, in which case the function is said to be a *point-wise discontinuous function*;

(*c*) everywhere-dense and closed, in which case the function is continuous;

(*d*) everywhere-dense in each interval of a set, and non-dense in each interval of another set external to the former one.

This last case (*d*) is not essentially distinct from the former ones.

By Hankel and others the term "totally discontinuous" has been applied to all functions which are neither continuous nor point-wise discontinuous.

187. It has been shewn by W. H. Young that a function can be constructed which is continuous at every point of any given inner limiting set of points, and is discontinuous at every other point of the interval.

Let E denote an inner limiting set, and let the function $f(x)$ be defined as follows:—

(1) At every point x of E, let $f(x) = x$.

(2) It has been shewn that a sequence of sets of non-overlapping intervals can be constructed such that the only points each of which is in an interval of every set are the points of E. Let Q be a limiting point of E which does not belong to E; then a number n exists such that Q is in an interval of the $n-1^{\text{th}}$ set, but not in one of the intervals of the n^{th} set. Let this interval of the $n-1^{\text{th}}$ set be of length d_Q; and at the point Q let $f(x) = x_Q + e^n d_Q$, where e is a fixed positive number less than unity; in the case $n=1$, we put $d_Q = e$.

(3) If R be a point which does not belong either to E or to its derivative, it must lie between two definite points A, B both of which belong to E or to E', and such that no point of E or of E' lies between A and B. If x_R be a rational number, let $f(x_R) = x_A$ or x_B, according as R is nearer to A or to B; when x_R is irrational, let $f(x_R) = x_R$; and if R be the middle point of AB, let $f(x_R) = x_R$.

* See a paper by W. H. Young, "Ueber die Eintheilung der unstetigen Funktionen und die Vertheilung ihrer Stetigkeitspunkte." *Wiener Sitzungsberichte*, vol. CXII. Abt. II a, 1903.

It is clear that the function so defined is discontinuous at every internal point of the interval AB, and at the end-point A it is continuous or discontinuous on the right, according as A does or does not belong to E; a similar result holds for B. It has thus been shewn that the function is discontinuous at every point which does not belong to E.

To shew that at any point P of E, the function is continuous, consider those intervals, one of each set in the sequence, which contain the point P: the lengths of these intervals will have a lower limit d which may be zero. In every interior point of d, we have $f(x)=x$; and thus, if P be interior to d, P is a point of continuity of the function. If P be an end-point of d, it is certainly continuous on the side towards the interval; and we have to shew that it is also continuous on the other side. Choose an arbitrarily small number σ, and an arbitrarily large integer m; then a number $n_1 > m$ can be found such that the n_1th, and all subsequent intervals of the sequence which contain P, are of length between d and $d+\sigma$. The piece of one of these intervals which is not a portion of d is of length $<\sigma$; and suppose that Q is a point in this piece which belongs to E' but not to E: then

$$|f(x_P)-f(x_Q)| = |x_P - x_Q - e^n d_Q| < |\sigma + e^m(d+\sigma)|$$

since $n \geqq n_1 > m$, and $d_Q < d+\sigma$. From this, there follows

$$|f(x_P)-f(x_Q)| < 2\sigma + e^m d < 3\sigma,$$

if m be chosen sufficiently great. If R be an interior point of an interval AB which contains in its interior no point of E' or of E, and if the points A, B be both so near to P that their distances from it are less than σ, we have $|f(x_P)-f(x_R)| < \sigma$, in virtue of the definition of $f(x_R)$; also if only one of the ends A, B be within the interval of length $<d+\sigma$, which has been chosen, then an interval further on in the sequence can always be found such that the middle point of AB is exterior to it, and thus the inequality $|f(x_P)-f(x_R)| < \sigma$, holds as in the former case. It has now been shewn that, for any arbitrarily chosen σ, a neighbourhood of P can be found such that for all points x in it, $|f(x_P)-f(x)| < 3\sigma$; therefore P is a point of continuity of the function. The case in which $d=0$, does not require separate treatment.

EXAMPLES.

1.* Let G denote a non-dense perfect set of points in the segment (0, 1), such that the end-points of the complementary intervals are rational points. Let $f(x)$ be defined thus:—at every irrational point inside an interval complementary to G, let $f(x)=x$; at every rational point of such interval, let $f(x)$ be equal to the value of x at the middle point of the interval; and at every point external to a complementary interval, let $f(x)=1$. This function is discontinuous except at the middle points of the intervals complementary

* See W. H. Young, *loc. cit.*

to G; thus the set of points of continuity is an enumerable set which contains no component dense-in-itself.

2.* With the same non-dense perfect set as in Ex. 1, let AB be a complementary interval of G, and M its middle point. At every rational point of AM except M, let $f(x)=x_A$, and at every rational point of MB except M, let $f(x)=x_B$; also at all points of (0, 1), except those for which the functional value has been already specified, let $f(x)=x$. In this case the points of continuity are non-dense and of the power of the continuum.

POINT-WISE DISCONTINUOUS FUNCTIONS.

188. A function being defined for the continuous domain (a, b), it can be shewn that, *if k be any fixed positive number, those points, at which the saltus of the function is $\geqq k$, form a closed set.*

This theorem follows immediately from the property of semi-continuous functions established in § 183, by considering the saltus-function. It may be proved directly as follows:—

If P be a limiting point of the set for which the saltus is $\geqq k$, then in any arbitrarily small neighbourhood of P there are points of the set; hence the fluctuation of the function in this neighbourhood is $\geqq k$, and therefore the saltus at P is $\geqq k$.

Moreover, such a limiting point P, of the set of points at which the saltus is $\geqq k$, must be a point of discontinuity of the second kind, at least on one side of P. If at P the function have a limit on the right, a neighbourhood PQ can be found such that the inner fluctuation in PQ is $<k$; hence inside PQ there can be no point at which the saltus is $\geqq k$; and therefore P is not a limiting point on the right, of the set for which the saltus is $\geqq k$. A similar remark applies to the left of P.

We have already defined a point-wise discontinuous function as one of which the points of continuity are everywhere-dense and unclosed in the domain of the function; this definition is that given by Dini†, and is equivalent to the following definition given by Hankel‡:—

A point-wise discontinuous function is one for which those points at which the saltus is $\geqq k$, an arbitrarily chosen positive number, form a non-dense set K, whatever value k may have.

That this set is closed has been shewn above.

To prove the equivalence of the two definitions, let it be assumed that in any arbitrarily chosen sub-interval (α, β), a point of continuity x_1 can be

* See W. H. Young, *loc. cit.*

† See *Grundlagen*, p. 81.

‡ *Math. Annalen*, vol. xx (1882), p. 90. This is a reproduction of Hankel's *Univ. Programm*, Tübingen, 1870, entitled "Untersuchungen über die unendlich oft oscillierenden und unstetigen Functionen."

found. A neighbourhood can be found for x_1, internal to (α, β), in which the fluctuation of the function is $< k$, and this neighbourhood can contain no point at which the saltus is $\geqq k$; hence the points at which the saltus is $\geqq k$ form a non-dense set K, since, interior to any sub-interval, a sub-interval can be found which contains no point of the set K.

Conversely, choose a descending sequence of values of k, say $k_1, k_2, k_3, \ldots$ which converges to zero, and let $K_1, K_2, K_3, \ldots$ be the corresponding non-dense closed sets, each of which necessarily contains the preceding one; then the set $M(K_1, K_2, K_3, \ldots)$ is the set of all the discontinuities of the function.

In accordance with § 89, this set is of the first category, and the complementary set, which is the set of points of continuity of the function, is everywhere-dense, and has the cardinal number of the continuum, being a set of the second category.

It will be observed that the set of all the points of discontinuity may be either everywhere-dense, or non-dense, in the whole or part of the domain of the variable. This set may be finite, enumerably infinite, or of the power of the continuum.

The set K, although non-dense, is not necessarily of content zero. By Harnack*, the term point-wise discontinuous function was only used for such functions as possess the property that the set K, for each value of k, has content zero. It will be seen that this latter case is of special importance in connection with the theory of integration.

It has been already shewn in § 186, that the points of continuity of the point-wise discontinuous function form an inner limiting set; and if

$$\{\delta_1\}, \{\delta_2\} \ldots \{\delta_n\} \ldots$$

be the sets of intervals complementary to the closed sets $K_1, K_2, \ldots K_n, \ldots$, they form a sequence of sets of non-overlapping intervals which define the set of points of continuity as their inner limiting set.

The whole theory of point-wise discontinuous functions is applicable to the case in which the domain of the variable is not a continuum, but is any perfect set. In this case also, the points of continuity of a point-wise discontinuous function are everywhere-dense relatively to the perfect domain, and the points at which the measure of discontinuity is $\geqq k$, form a closed set, non-dense relatively to the domain of the variable. That this is the case may be shewn by making the points of the perfect set correspond in order to the points of a continuous interval, as explained in § 75. The points of discontinuity, and those of continuity, relatively to the perfect domain, are sets of the first and the second category respectively, relative to that domain.

* *Math. Annalen*, vol. XIX, 1882, p. 242, and vol. XXIV, 1884, p. 218.

189. The domain of the variable being either a continuous interval or any perfect set, let us suppose that at every point the oscillation (see § 181) of the function both on the right and on the left is $< k$; there can then only be a finite number of points at which the saltus is $\geqq k$; *i.e.* the set K is finite.

For if K were not finite, it must contain a limiting point P, which has been shewn in § 188 to be a point of discontinuity of the second kind. Any arbitrarily small neighbourhood of P, on one side at least, must therefore contain points at which the saltus is $\geqq k$, and hence the oscillation at P on this side could not be $< k$.

The domain can therefore be divided into a finite number of parts within each of which there is no point at which the saltus is $\geqq k$; and it follows that the domain can be divided into a finite number of parts, within each of which the fluctuation of the function is $< k$.

If, at each point of a set which is everywhere-dense in the domain of the variable, there exist a limit of the function on one side at least, then the function is either point-wise discontinuous, or else it is continuous.

In any interval, containing points of the domain, a point can be found which has a neighbourhood on one side at least in which the inner fluctuation is $< k$; within such neighbourhood the saltus is everywhere $< k$; hence the points of K are non-dense in the domain, and thus the function is either point-wise discontinuous, or else it is continuous.

A particular case of this theorem is that *a function defined for a continuous interval, and having ordinary discontinuities only, is point-wise discontinuous.*

Among such functions, the *monotone* functions form an important class. A *monotone function* is one such that for every pair of values x_1, x_2 of the variable, such that $x_2 > x_1$, the condition $f(x_2) \geqq f(x_1)$ is satisfied; or else, for every such pair, the condition $f(x_2) \leqq f(x_1)$ is satisfied. Since there can be no oscillations in the neighbourhood of any point, every discontinuity must be an ordinary one. It follows that *every monotone function is either point-wise discontinuous, or else continuous.*

If the function be defined for a continuous interval, and all the points of discontinuity be ordinary ones at least on one side, then the set K of points, at which the saltus is $\geqq k$, is a set of content zero.

The set K can be resolved into a perfect set G and an enumerable set; the set G contains points which are limiting points on both sides, and at such a point the oscillation both on the right and on the left must be $\geqq k$; it follows that the set G is non-existent, and that K is therefore an enumerable closed set, which has necessarily content zero.

The theorem still holds if there be points of discontinuity of the second kind which form a set of content zero, for these points may be enclosed in a finite number of intervals whose sum is arbitrarily small; the theorem can then be applied to each of the remaining intervals of the domain.

190. At a point P at which a given point-wise discontinuous function is discontinuous, let $x_1, x_2, x_3, \ldots x_n, \ldots$ be any sequence of points converging to P, and let us suppose that $f(x_1), f(x_2), \ldots f(x_n), \ldots$, the set of values of the given function at the points $\{x_n\}$, converges to a limit U; the value of U will depend upon the choice of the particular sequence, subject to the condition of convergence of $\{f(x_n)\}$. If P be a point of discontinuity of the first kind, U is capable of having two values only, viz. $f(x+0)$ and $f(x-0)$; but if the discontinuity of the function at P be of the second kind, U may have any one of the values $\overline{f(x+0)}$, $\underline{f(x+0)}$, $\overline{f(x-0)}$, $\underline{f(x-0)}$, or it may possibly have other values lying between the greatest and least of these four. The possible nature of the set of all values of U at the point x will now be investigated. This set is the aggregate of functional limits at x (see § 178).

In the first place* *the set of all values of U at a point P is a closed set.*

For let $U_1, U_2, U_3, \ldots U_n, \ldots$ be a convergent sequence of values of U, of which U_ω is the limit; it will be shewn that U_ω itself belongs to the set of values of U, and thus that this set is closed.

Let U_r be the limit of the convergent sequence

$$f(x_1^{(r)}),\quad f(x_2^{(r)}),\ \ldots f(x_n^{(r)})\ldots;$$

we may choose r so great that $|U_\omega - U_r| < \frac{1}{2}\epsilon$, for this and all greater values of r. We can then choose n such that $|U_r - f(x_n^{(r)})| < \frac{1}{2}\epsilon$; it follows that, for sufficiently great values of r and n,

$$|U_\omega - f(x_n^{(r)})| < \epsilon,$$

and as the positive number ϵ is arbitrarily small, the theorem is established; in fact a sequence $\{f(x_n^{(r)})\}$ can be found which converges to U_ω, whilst $\{x_n^{(r)}\}$ converges to x.

A point-wise discontinuous function can be constructed so that, at a point x_0, the set of values of U may be any prescribed closed set G.*

To establish this theorem, we observe that if G be unenumerable it may be replaced by an enumerable set G_1 everywhere-dense in G. Let

$$V_1,\ V_2,\ V_3, \ldots$$

denote the set G_1, arranged in the order-type ω.

Next, choose an enumerable sequence $x_1, x_2, \ldots x_n, \ldots$ of values of x having the single limiting point x_0; and arrange the sequence $\{x_n\}$ in the order-type ω^2. The sequence $\{x_n\}$ may thus be split up into an enumerable set of sequences $\{x_{1r}\}, \{x_{2r}\}, \ldots \{x_{sr}\}, \ldots$ each of which has the limit x_0. The

* These theorems were given by Bettazzi, see *Rendiconti di Palermo*, vol. VI, p. 173.

function $f(x)$ may be defined by the specifications, that $f(x)=0$, for all values of x which do not belong to the sets $\{x_{1r}\}$, $\{x_{2r}\}$, ...; and that

$$f(x_{sr})=v_{sr}, \text{ where } v_{s1}, v_{s2}, \dots v_{sr}, \dots$$

is a sequence chosen so as to converge to the limit V_s. The function $f(x)$ so defined is continuous at every point except $x_0, x_1, \dots x_n, \dots$, and it has the required property; since the points of G_1, and therefore of G, are all values of U at the point x_0.

EXAMPLES.

1.* If $f(x)$, $\phi(x)$ be two point-wise discontinuous functions defined for the same interval, there is an everywhere-dense set of points at each of which both functions are continuous. This theorem follows at once from the fact that the points common to two sets of the second category also form a set of the second category, and that this holds for every sub-interval contained in the given interval.

2. Let (x) denote the positive or negative excess of x above the integer nearest to it, and if x be half-way between two successive integers, let $(x)=0$. Let a function† $f(x)$ be defined for the interval (0, 1) as the limit of

$$\frac{(x)}{1}+\frac{(2x)}{4}+\frac{(3x)}{9}+\dots+\frac{(nx)}{n^2},$$

when n is indefinitely increased. The function $f(x)$ is a point-wise discontinuous function, in which the set K of points, at which the saltus is $\geq k$, is finite for each positive value of k. It can be proved that, if $x=m/2n$, where m and $2n$ are relative primes, then

$$f\left(\frac{m}{2n}+0\right)=f\left(\frac{m}{2n}\right)-\frac{\pi^2}{16n^2}, \quad f\left(\frac{m}{2n}-0\right)=f\left(\frac{m}{2n}\right)+\frac{\pi^2}{16n^2}.$$

For values of x not of the above form, $f(x)$ is continuous. The number of points of K is the number of irreducible proper fractions having even denominators $2n$, such that $\pi^2/8n^2 \geq k$. The set of all the points of discontinuity is everywhere-dense in the interval (0, 1).

3. Let‡ $y=c$, for all rational values of x; and $y=d$, for all irrational values of x. This function is totally discontinuous.

4. Let§ $f(x)=1$, for all values of x in the interval (0, 1), except $x=\frac{1}{2^n}$, $(n=1, 2, 3, \dots)$, for which $f(x)=0$. At each of the points $\left(\frac{1}{2^n}\right)$ there is a saltus equal to unity. This function is point-wise discontinuous, and the content of K is zero, for every value of k.

5. In§ the interval $\left(\frac{1}{2}, 1\right)$ of x, let $f(x)=1$; in the interval $\left(\frac{1}{2^2}, \frac{1}{2}\right)$, let $f(x)=\frac{1}{2}$; and in general, in the interval $\left(\frac{1}{2^{n+1}}, \frac{1}{2^n}\right)$, let $f(x)=\frac{1}{2^n}$. In this case the point-wise discontinuous function $f(x)$ is such that the number of points at which the saltus is $\geq k$, is finite for every value of $k>0$.

* See Volterra, *Giornale di Mat.* vol. XIX, 1881.

† See Riemann's *Ges. Werke*, p. 242.

‡ Dirichlet's *Werke*, p. 132.

§ Hankel, *Math. Annalen*, vol. XX.

6. The* points of a continuous interval (0, 1) may be put into correspondence with the points of a non-dense set of points, dense-in-itself, contained in an interval (a, b), in such a manner that the relative order of two points of the interval (0, 1) is the same as that of the corresponding points in (a, b). Such a correspondence is defined by a point-wise discontinuous monotone function $y=f(x)$.

7.† Let the numbers of the interval (0, 1) be expressed as finite or infinite decimals $x=\cdot a_1a_2a_3\ldots a_n,\ldots$ and let $f(x)=\left(\frac{a_1}{10}\right)^2+\left(\frac{a_2}{100}\right)^2+\ldots$. The function $f(x)$ is monotone, and is discontinuous for every value of x represented by a finite decimal. The set of points K for a given value of k is finite. The function $f(x)$ defined by $f(x)=\cdot 0a_1 0a_2 0a_3\ldots$ has similar properties.

8.‡ Let the points of the interval (0, 1) be represented by decimals, and consider the set G_0 of those points for which only the digits 0 and 1 occur in the decimal representation, excluding those points for which all the figures are 0, from and after some fixed place. The set G_0 is non-dense in the interval (0, 1), and has the cardinal number c. Any point x_0 of G_0 is represented by $\cdot a_1a_2a_3\ldots a_n\ldots$, where a_n is 0 or 1. Let ξ be a fixed point $\cdot b_1b_2\ldots b_n\ldots$ of G_0, and let x_ξ denote $x_0+2\xi\equiv\cdot c_1c_2\ldots c_n\ldots$; so that $c_n=a_n+2b_n$. With ξ fixed, let the set of all points x_ξ be denoted by G_ξ; the points of G_ξ are all different from those of G_0, and for two values ξ, ξ' of ξ, the sets G_ξ, $G_{\xi'}$ have no point in common. For two numbers $\cdot c_1c_2\ldots c_n\ldots$, $\cdot c_1'c_2'\ldots c_n'\ldots$ are identical only when $c_n=c_n'$, which holds only when $a_n=a_n'$, and $b_n=b_n'$. If we read off in the dyad scale the decimal representation of ξ, we obtain, by giving ξ all the values in G_0, every point in (0, 1) except the point 0, and these once only; let the point which, by thus using the dyad scale, corresponds to ξ be denoted by (ξ). Now let $f(x)$ be defined by the rules $f(x_\xi)=(\xi)$, $f(x_0)=0$, and $f(x)=0$ for all other values of x. The point-wise discontinuous function $f(x)$ so defined for the interval (0, 1) is such that at all the points of the unenumerable set G_ξ the saltus is (ξ); the set of all the points of discontinuity is non-dense in (0, 1); and $f(x)$ is constant, and $=0$, in an everywhere-dense set of linear intervals.

9.‡ Let the points x of the interval (0, 1) be expressed as radix-fractions in the scale of 3. Let G_0 be the set of points for which all the figures of the radix-fraction are 0 and 1, excepting those points for which all the figures are 0 after some fixed place. Let G_n consist of all the points which contain the digit 2 in at most the first n places, but are also such that the nth figure is 2; then G_n is non-dense, and of cardinal number c. There are left only those points for which the radix-fractions contain the digit 2 an infinite number of times; and these points belong to a set H for which the radix-fractions contain an infinite number of digits other than 2, or to a set G for each point of which every digit is 2, from and after some fixed one. Each point of G can be represented by a terminating radix-fraction which contains only a finite number of 2's, and can be added to a G_n. Let $G_0, G_1, G_2, \ldots$, when so increased, become $\bar{G}_0, \bar{G}_1, \bar{G}_2\ldots$; and take a sequence of decreasing numbers $g_0, g_1, g_2, \ldots$. Let the function $f(x)$ be defined by the rules $f(x)=g_n$, if x is a point of $\bar{G}_n$, and $f(x)=0$ for all points of H. The point-wise discontinuous function $f(x)$ is continuous at all the points of H, and the points of discontinuity are everywhere-dense in (0, 1), and of cardinal number c.

* Harnack, *Math. Annalen*, vol. XXIII. † Peano, *Riv. di Mat.* vol. I.
‡ Schönflies, *Göttinger Nachrichten*, 1899.

DEFINITION OF POINT-WISE DISCONTINUOUS FUNCTIONS BY EXTENSION.

191. Let us suppose a function $f(x)$ to be defined for a domain which consists of a set of points which is dense-in-itself but not closed, and further let us assume that $f(x)$ is continuous in this domain. The new domain obtained by adding to the original domain those of its limiting points which do not belong to it may be spoken of as the *extended domain*. It has been pointed out in § 190 that, at a point a of the extended domain, which does not belong to the original domain, there is an aggregate of functional limits which is certainly a closed set, and may consist of a finite, or an infinite, set of numbers.

Let us now define a function $\phi(x)$, for the extended domain, in the following manner:—At each point of the original domain, which may be called a *primary point*, let $\phi(x) = f(x)$; at each point a, which may be called a *secondary point*, and which does not belong to the original domain, attribute to $\phi(x)$ the values contained in the aggregate of functional limits of $f(x)$ at a; this function $\phi(x)$ may then be multiple-valued at any secondary point. The new function $\phi(x)$ defined for the extended domain may be spoken of as *the function obtained by extension of* $f(x)$; and those points for which $\phi(x)$ is multiple-valued are regarded as points of discontinuity at which the measure of discontinuity is the excess of the greatest over the least value of the function at the point.

It will be shewn that *the extended function* $\phi(x)$ *is point-wise discontinuous in the extended domain, unless it be continuous*. This gives rise to a method of constructing point-wise discontinuous functions which has been employed by Brodén in various special cases. Since we may so choose the original domain that it shall consist of an enumerable set of points, the method includes one for the construction of a point-wise discontinuous function from an enumerable set of specifications.

To prove that the extended function $\phi(x)$ is at most point-wise discontinuous, it is sufficient to shew that $\phi(x)$ is continuous at all points of the original domain G, which is a set that is everywhere-dense in the extended domain G'.

Let x be a point of G, and let it be the limiting point of a convergent sequence $(x_1', x_2', x_3', \ldots)$ of which all the points belong to G'. Consider the aggregate $\{\phi(x_1'), \phi(x_2'), \ldots\}$, where $\phi(x_1'), \phi(x_2'), \ldots$ have any of the values which belong to the points $x_1', x_2', x_3', \ldots$. Now a point x_n of G can be found such that $|x_n' - x_n| < \eta_n$, and $|\phi(x_n') - f(x_n)| < \epsilon_n$, where η_n, ϵ_n are independent arbitrarily small numbers. If we take a sequence of values of η, such that $\eta_1 > \eta_2 > \eta_3, \ldots$, with zero as its limit, and also a similar sequence of the ϵ numbers, then the sequence $(x_1, x_2, x_3, \ldots)$ has the same limit x as the sequence $(x_1', x_2', x_3', \ldots)$, and the aggregate $\{\phi(x_1'), \phi(x_2'), \ldots\}$ has the same limit as the convergent aggregate $\{f(x_1), f(x_2), \ldots\}$ viz. $f(x)$ or $\phi(x)$; and thus the theorem is established.

It will be observed that the values of $\phi(x)$ at all the secondary points in an arbitrarily small neighbourhood of a secondary point α depend only on the values of $f(x)$ in that same neighbourhood; it follows therefore that α is a point of continuity or of discontinuity of $\phi(x)$ according as the aggregate of functional limits of $f(x)$ at α consists of one number or of more. In the latter case the measure of discontinuity of $\phi(x)$ at α is the excess of the greatest over the least of the numbers belonging to the values of $\phi(x)$ at the point.

It can be shewn that a point-wise discontinuous function can be so constructed that, at a given secondary point, the values of the function may be an arbitrarily assigned closed set.

192. Although a class of point-wise discontinuous functions may be obtained by extension of a continuous function defined for a primary domain, dense-in-itself but unclosed, yet not every point-wise discontinuous function can be generated in this manner.

Let $f(x)$ be a point-wise discontinuous function in a domain which is either a continuum or a perfect set of points.

Consider the function $\phi(x)$ obtained by taking the values of $f(x)$ as given only at its points of continuity, and extending this function to the complete domain, in the manner explained above.

At each point of discontinuity of $f(x)$ there is a saltus k_f, and at that point the function $\phi(x)$, obtained by extending the set of values of $f(x)$ at its point of continuity, has a measure of discontinuity k_ϕ, which will be zero in case $\phi(x)$ be continuous at the point; but in any case the condition $k_\phi \leqq k_f$ is satisfied, since, within any neighbourhood of the point, the fluctuation of $\phi(x)$ cannot be greater than that of $f(x)$.

If $k_\phi = 0$ at any point of discontinuity of $f(x)$, that point may be said to be a point of *unessential* discontinuity of the function $f(x)$; and if $k_\phi > 0$, the point is one of *essential* discontinuity.

Let now a function $\chi(x)$ be defined for the whole domain as follows:— At every point of continuity of $f(x)$, and at every point of discontinuity at which $k_f = k_\phi$, let $\chi(x) = 0$; at each point at which $k_f > k_\phi$, let

$$\chi(x) = k_f - k_\phi.$$

The function $\chi(x)$ is not necessarily continuous at every point at which it is zero. At a point x_1 at which $\phi(x)$ is continuous, the measure of discontinuity of $\chi(x)$ is k_f, or $\chi(x_1)$; but this is not necessarily the case if $\phi(x)$ be not continuous at x_1. This function $\chi(x)$ may be called a *point-wise discontinuous null-function*.

By subtracting from $f(x)$ a function $\psi_1(x)$ which never exceeds, at any point x, in absolute value, the value of $\chi(x)$, we obtain a function $\phi_1(x)$ of which the measure of discontinuity is everywhere $= k_\phi(x)$.

The function* $\phi_1(x)$ may be spoken of as the *most nearly continuous function associated with* $f(x)$.

It thus appears that a point-wise discontinuous function can always be expressed as the sum of a point-wise discontinuous null-function and the most nearly continuous function associated with the given function.

The latter function $\phi_1(x)$ has only those discontinuities which necessarily arise from the values of the given function at its points of continuity, and is independent of the parts of the discontinuities which arise out of the functional values of $f(x)$ at the points of discontinuity. The null-function depends upon the unessential parts of the discontinuity of $f(x)$.

EXAMPLES.

1.† Let $f(x)=0$, for $x=0, \frac{1}{\pi}, \frac{1}{2\pi}, \frac{1}{3\pi}, \ldots$, and for all other positive and negative values of x, let $f(x)=\cos\frac{1}{x}$. The function $\phi(x)$ associated with $f(x)$ agrees with $\cos\frac{1}{x}$ at every point except $x=0$, where $\phi(x)$ is represented by $(-1, +1)$. The measure of discontinuity k_f is zero except at $\frac{1}{\pi}, \frac{1}{2\pi}, \frac{1}{3\pi}, \ldots$, where $k_f=1$, and at $x=0$, where $k_f=2$; the measure k_ϕ vanishes everywhere except at $x=0$, where $k_\phi=2$. The function $\chi(x)$ vanishes except at $\frac{1}{\pi}, \frac{1}{2\pi}, \ldots$, where it is 1; it vanishes at $x=0$, but is discontinuous at that point.

2.† Let $f(x)$ vanish except at the points $x=\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots \frac{1}{n}, \ldots$, where $f(x)=1$. The function $\phi(x)$ is everywhere zero, and thus k_ϕ is everywhere zero. The function $\phi_1(x)$ is everywhere zero, and k_f is zero except at $0, \frac{1}{2}, \frac{1}{3}, \ldots \frac{1}{n}, \ldots$, where $k_f=1$. In this case $f(x)-\phi_1(x)=f(x)$.

3.† A point-wise discontinuous function $f(x)$ can be constructed‡ such that the function $\phi(x)$ may have at a point x_0, the values belonging to a prescribed closed set G, in accordance with Bettazzi's theorem (§ 190). If G be unenumerable, choose an enumerable set G_1 dense in G, and let $g_1, g_2, g_3, \ldots$ be the points of G_1. Take a set of intervals $\{\delta_n\}$, where δ_n is $\left(x_0+\frac{1}{2^n}, x_0+\frac{1}{2^{n-1}}\right)$, and define $f(x)$ as follows:—in $\delta_1, \delta_3, \delta_5, \delta_7, \ldots$ let $f(x)=g_1$; in $\delta_2, \delta_6, \delta_{10}, \ldots$ let $f(x)=g_2$; in $\delta_4, \delta_{12}, \delta_{20}, \ldots$ let $f(x)=g_3$; in general $f(x)=g_n$,

* This definition is not in complete agreement with that of Schönflies, see *Bericht*, p. 134, to whom the term is due. Some erroneous statements of Schönflies, in this connection, were pointed out and corrected by Hahn; see *Monatshefte f. Math.* vol. XVI, 1905.

† See Hahn, *loc. cit.*

‡ This is contrary to a statement of Schönflies, see *Bericht*, p. 135.

in the first free interval, and in every second of the following free intervals; further let $f(x)=g_1$, for $x \leqq x_0$. The function $f(x)$ is point-wise discontinuous, the points of discontinuity being x_0, and the points $x_0+\frac{1}{2^n}$. The function $\phi(x)$ has two values at the points $x_0+\frac{1}{2^n}$; and at x_0 it has all the values of G_1, and therefore all those of G.

FUNCTIONS WITH LIMITED TOTAL FLUCTUATION.

193. Let a function $f(x)$ be defined for the continuous interval (a, b). Suppose the interval (a, b) to be divided into a number n_1 of non-overlapping sub-intervals the greatest of which is d_1; let these sub-intervals be divided into smaller ones, so that the total number of sub-intervals is now n_2, and the greatest of them is d_2. Proceed in this manner to continually sub-divide the sub-intervals according to some prescribed law, so chosen that the numbers $n_1, n_2, n_3, \ldots$ form a sequence of continually increasing numbers, and that the numbers $d_1, d_2, d_3, \ldots$ form a convergent sequence converging to the limit zero. Such a system of indefinitely continued sub-divisions of the interval (a, b) may be spoken of as a *convergent system of sub-intervals.*

If the function $f(x)$ be such that, any particular convergent system of sub-intervals of (a, b) being taken, and the sum of the fluctuations of $f(x)$ in the n_r sub-intervals into which (a, b) is divided at the rth stage of the process of successive sub-division being denoted by $\sum_{m=1}^{n_r} \Delta_{r,m}$, this sum is for every value of r less than some fixed finite number, then $f(x)$ is said to be a function with limited total fluctuation in the interval (a, b).

It will be shewn that, when the condition stated in this definition is satisfied, then *the sums $\sum_{m=1}^{n_r} \Delta_{r,m}$ have an upper limit L, which may be called the total fluctuation of the function in the interval (a, b), for the prescribed convergent system of sub-intervals.*

For, if an interval (α, β) be divided into two parts (α, γ), and (γ, β), by a point γ, it is clear that the fluctuation in (α, β) is not greater than the sum of the fluctuations in (α, γ) and (γ, β); and therefore, when (α, β) is divided into any number of parts, the sum of the fluctuations in those parts is greater than, or equal to, the fluctuation in (α, β). It thus appears that, for the given succession of sub-intervals, the numbers $\sum_{m=1}^{m=n_r} \Delta_{r,m}$, which are essentially positive, never diminish as n increases. Therefore, since they are all less than some fixed number, they have a fixed finite limit L.

If $f(x)$ have a limited total fluctuation L for a prescribed convergent system of sub-intervals, then it has also a limited total fluctuation for any other such convergent system of sub-intervals. Further, the total fluctuation L, has an*

* See Study, *Math. Annalen*, vol. XLVII, p. 299.

upper boundary M, and a lower boundary $\mu \geqq \frac{1}{2}M$, when all possible convergent systems of sub-intervals are considered.

Let $\Delta_1, \Delta_2, \ldots \Delta_m$ denote the fluctuations in those sub-intervals $\delta_1, \delta_2, \ldots \delta_m$, which, at any stage, belong to the originally prescribed mode of sub-division, and let $\Delta_1', \Delta_2', \ldots \Delta'_{m'}$ be the fluctuations in the sub-intervals $\delta_1', \delta_2' \ldots \delta'_{m'}$, at some stage of any other mode of sub-division of the interval. If η be a positive number smaller than the smallest of the numbers δ', we may take m to be so large that none of the intervals δ is greater than η. Now let us suppose the two sets of sub-intervals to be superimposed; then any interval δ_r will be divided, by means of the points of division in the second set, into not more than two parts δ_{r1}'', δ_{r2}'', with fluctuations Δ_{r1}'', Δ_{r2}'', if divided at all; we have then

$$\Delta_r \leqq \Delta_{r1}'' + \Delta_{r2}'' \leqq 2\Delta_r.$$

Now every δ' interval is made up of δ intervals and δ'' intervals; therefore $\Sigma\Delta' \leqq \Sigma\Delta'' + S$, where S denotes the sum of the fluctuations in the undivided δ intervals. It follows that $\Sigma\Delta' \leqq 2\Sigma\Delta \leqq 2L$. It thus appears that the numbers $\Sigma\Delta'$, corresponding to any arbitrarily prescribed system of sub-divisions, are all not greater than $2L$; and $\Sigma\Delta'$ has therefore, as in the case of the original system, an upper limit L', and this is $\leqq 2L$. For every possible system of sub-divisions, the numbers L' form an aggregate of positive numbers which do not exceed $2L$; and they therefore have an upper boundary M, and a lower boundary μ. Moreover, a system of successive sub-divisions can be defined such that the limit of the sum of the fluctuations is μ, in case the lower boundary μ be attained by the set of numbers L', or is $\mu + \epsilon$, where ϵ is less than an arbitrarily chosen number, in case μ be not attained. We may therefore take $L = \mu$, or $\mu + \epsilon$, as the case may be; and then, for any system of sub-divisions δ', $\Sigma\Delta' \leqq 2\mu$, or $2(\mu + \epsilon)$: it thus follows that the upper boundary M is not greater than 2μ, or that $\mu \geqq \frac{1}{2}M$. The numbers M and μ are the limits of indeterminacy of the total fluctuation of the function in (a, b), where M cannot fall outside the interval $(\mu, 2\mu)$.

194. A function with limited total fluctuation can have no points of discontinuity of the second kind, and thus for such a function the limits $f(x+0), f(x-0)$ on the right and on the left must both exist at every point; except that at the points a, b, only the limits $f(a+0)$, $f(b-0)$ can exist. For, in any arbitrarily small neighbourhood of a point of discontinuity of the second kind, the function makes an infinite number of oscillations which are greater than some fixed finite number; and thus, in such a neighbourhood, the total fluctuation of the function cannot be finite.

It is clear that, in a function with limited total fluctuation, the sums of the saltuses on the right and on the left

$$\Sigma\,|f(x+0) - f(x)|, \quad \Sigma\,|f(x-0) - f(x)|,$$

for all the points of discontinuity, are finite. It follows that the points of discontinuity form an enumerable set, since there can be only a finite number at which either saltus exceeds any arbitrarily chosen positive number. A function of this kind may however have oscillations in every sub-interval, but it is necessarily either point-wise discontinuous or else continuous.

The following theorem will now be established:—

If a function with limited total fluctuation have at no point an external saltus, then $M = \mu$; and thus the total fluctuation of the function is the same for every convergent system of sub-divisions.

For let us assume that, if possible, there exists a convergent system of sub-divisions, such that $\sum_{m=1}^{n_r} \Delta_{r,m}$ converges to a value G less than M. The interval (a, b) can also, by hypothesis, be divided into a number n of parts, such that the sum $\Sigma\Delta'$ of the fluctuations in those parts is greater than G, say $= G + \alpha$. Let this sub-division into n parts be superimposed on the set of sub-divisions for which $\sum_{m=1}^{n_r} \Delta_{r,m}$ is the sum of the fluctuations; we may assume that r is so great that not more than one of the $n-1$ points of the former sub-division is in any one of the intervals of the latter set, while none of the end-points of the two sets coincide except at a and b. Let us suppose that x is one of these $n-1$ points, and that it falls in the interval δ for which $\Delta_{r,m}$ is the fluctuation, dividing it into two parts, δ_1' and δ_2'. Assuming that there is no external saltus, if η be any fixed positive number, we know that if δ_1', δ_2' are sufficiently small, then

$$\Delta_{r,m} = |f(x+0) - f(x-0)| + \sigma_1, \text{ where } \sigma_1 < \eta\,;$$

moreover, under a similar condition, the fluctuations Δ_1', Δ_2' in δ_1', δ_2' are such that

$$\Delta_1' = |f(x) - f(x-0)| + \sigma_2, \text{ where } \sigma_2 < \eta$$

and

$$\Delta_2' = |f(x) - f(x+0)| + \sigma_3, \text{ where } \sigma_3 < \eta.$$

Now r may be chosen so great, and consequently d_r so small, that these conditions are satisfied for each of the $n-1$ intervals of the originally assumed set which contain points x. We have then $\Delta_1' + \Delta_2' - \Delta_{r,m} = \sigma_2 + \sigma_3 - \sigma_1$, provided $f(x)$ lies between $f(x+0)$ and $f(x-0)$; hence the sum of the differences $\Delta_1' + \Delta_2' - \Delta_{r,m}$ taken for all those intervals which contain one of the $n-1$ points x is less than $2(n-1)\eta$; and it cannot be negative. It follows that the sum of the fluctuations in the intervals obtained by superimposing on the sub-divisions, for which the $\Delta_{r,m}$ are the fluctuations, the $n-1$ points x, is $< G + 2(n-1)\eta$. But the sum of these fluctuations is certainly $\geqq G + \alpha$; and since η is arbitrarily small, and independent of n, these two relations are incompatible with one another, *i.e.* α must be zero. It has thus been shewn that, provided there be no external saltus at any point, G must equal M, and this therefore is the limit of $\Sigma\Delta$ for every convergent system of sub-divisions of (a, b).

In case there be points at which there is an external saltus, the preceding proof can still be applied to shew that the total fluctuation is the same for any two convergent systems of sub-divisions of the interval (a, b), provided no point at which there is an external saltus be an end-point of an interval in either system. Moreover, in case there be an external saltus at the point x, the value of $\Delta_{r,m}$ is either $|f(x)-f(x+0)|+\sigma_1$ or $|f(x)-f(x-0)|+\sigma_1$; and we see that $\Delta_1'+\Delta_2'-\Delta_{r,m}=\sigma_2+\sigma_3-\sigma_1+s(x)$, where $s(x)$ denotes the external saltus at x, and is equal to the smaller of the two numbers

$$|f(x)-f(x+0)|, \quad |f(x)-f(x-0)|.$$

If we take the $n-1$ points x to consist of all those points at which there is an external saltus greater than some fixed number β, we see that the sum of the fluctuations in the set of intervals, obtained by superimposing the $n-1$ points x on the system for which $\Sigma\Delta_{r,m}$ is the sum of the fluctuations, is $\Sigma\Delta_{r,m}+S_\beta+\gamma$, where S_β denotes the sum of those external saltuses, all of which are greater than β, and γ is the sum of the $n-1$ numbers $\sigma_2+\sigma_3-\sigma_1$, and is therefore arbitrarily small. It thus appears that the total fluctuation for a convergent system of sub-intervals, such that no point, at which there is an external saltus, is ever an end-point, must be μ; whereas if those points at which the external saltus is greater than β be end-points of intervals, the total fluctuation is $\mu+S_\beta$. If a sequence of descending values be given to β, the sum S_β converges to a fixed finite number, which is the sum or limiting sum of all the external saltuses. Thus we obtain the following theorem:—

If a function with limited total fluctuation be such that there are points at which the function has an external saltus, then the difference $M-\mu$ between the upper and lower boundaries of the total fluctuation L of the function, for convergent systems of sub-intervals, is equal to the sum of all the external saltuses. If a convergent system of sub-intervals be such that no point at which there is an external saltus is an end-point of any sub-interval, then for such a system $L=\mu$. If, however, the system be such that every point at which there is an external saltus is an end-point of an interval of the system, then $L=M$.

It thus appears that, provided $f(x)$ be at every point intermediate in value between $f(x+0)$ and $f(x-0)$, the total fluctuation in (a, b), for a function of the class considered, is a definite number. This number is unaltered by changing the value of $f(x)$ at a point of discontinuity of the function, provided no external saltus be introduced. For example, we may assign to $f(x)$ the value $\frac{1}{2}\{f(x+0)+f(x-0)\}$ at every point, without altering the total fluctuation of the function.

If a function $f(x)$, for which an external saltus exists at points of a certain set, be replaced by a new function $\phi(x)$, differing from $f(x)$ only at the points of the set, and such that $\phi(x)$ has nowhere an external saltus, then the new function $\phi(x)$ has its total fluctuation equal to a definite number, which is independent of the particular system of sub-intervals employed. Consider a system of sub-intervals for which no point of the set is ever an end-point of a sub-interval; then the limit of the sum of the fluctuations of $f(x)$ for this system is μ. If at each point of the set we remove the external saltus, by there substituting $\phi(x)$ for $f(x)$, we diminish this minimum total fluctuation μ of $f(x)$ by the sum of the external saltuses, and this is $M-\mu$. It thus appears that the total fluctuation of the function $\phi(x)$ is $2\mu - M$, and this is a definite number independent of any particular system of sub-intervals.

195. A function $\phi(x)$ defined for the interval (a, b), and such that for every pair of points x_1, x_2 in the interval, for which $x_2 > x_1$, the condition $\phi(x_2) \geqq \phi(x_1)$ is satisfied, has been defined, in § 189, to be *monotone* in (a, b). If for every such pair of points the condition $\phi(x_2) \leqq \phi(x_1)$ is satisfied, $\phi(x)$ is also monotone. In the former case $\phi(x)$ never diminishes, and in the latter case it never increases, as x is increased through the interval from a to b.

A function with limited total fluctuation can always be expressed as the difference of two functions, each of which is monotone in the interval for which the function is defined, and neither of which diminishes as the variable increases.

The importance of the class of functions with limited total fluctuation, in connection with the theory of Fourier's and other series, depends upon their possession of this property.

To prove the theorem, let the upper boundary of the total fluctuation of the function, for the interval (a, x), be denoted by $M_a{}^x$. We then see that $f(x+h) - f(x) \leqq M_x{}^{x+h} \leqq M_a{}^{x+h} - M_a{}^x$; it follows that $M_a{}^x - f(x)$ is a monotone function which never diminishes as x is increased. The function $M_a{}^x + f(x)$ has the same property, since $f(x+h) - f(x) \geqq -M_x{}^{x+h}$. Hence, if

$$\phi_1(x) = \tfrac{1}{2}\{M_a{}^x + f(x)\}, \quad \phi_2(x) = \tfrac{1}{2}\{M_a{}^x - f(x)\},$$

we can express $f(x)$ in the form $\phi_1(x) - \phi_2(x)$, where $\phi_1(x)$, $\phi_2(x)$ are both monotone non-diminishing functions as x increases through the interval (a, b). The converse property is easily seen to hold, that every function expressible in this manner has a limited total fluctuation. For the fluctuation of a monotone function in an interval is the difference of the functional values at the ends of the interval.

196. The class of functions of which the properties have been investigated above has been defined in a different manner by Jordan*, and it has been shewn by Study† that the two definitions are completely equivalent to one another.

Let us denote by $D_{r,m}$, the absolute difference of the functional values at the end-points of that sub-interval belonging to a convergent system of sub-

* *Cours d'Analyse*, vol. I, p. 55.

† *Math. Annalen*, vol. XLVII, p. 55.

intervals, the fluctuation in which has been denoted above by $\Delta_{r,m}$. Let us consider the sum $\sum_{m=1}^{n_r} D_{r,m}$, which is equivalent to

$$|f(a)-f(x_{r,1})|+|f(x_{r,1})-f(x_{r,2})|+\ldots+|f(x_{r,m-1})-f(x_{r,m})| + \ldots + |f(x_{r,n_r-1})-f(b)|,$$

where $a, x_{r,1}, x_{r,2}, \ldots, b$ are the points of the sub-intervals at the rth stage of the process of successive sub-division of (a, b). If the numbers $\sum_{m=1}^{n_r} D_{r,m}$, for every value of r, and for every possible convergent system of sub-divisions of (a, b) be all less than some fixed finite number, then the function is said to be a function with *limited total variation* (à variation bornée) in (a, b); and the upper limit of the numbers $\sum_{m=1}^{n_r} D_{r,m}$ for a particular system of sub-divisions, as r is indefinitely increased, is said to be the *total variation* of $f(x)$ in (a, b) for that particular convergent system of sub-divisions.

It will be shewn that a function with limited total fluctuation is also a function with limited total variation; and the converse. The first part of this theorem follows at once from the fact that, in any sub-interval, $D_{r,m} \leqq \Delta_{r,m}$, and therefore $\sum_{m=1}^{n_r} D_{r,m}$ is certainly less than a fixed finite number, if $\sum_{m=1}^{n_r} \Delta_{r,m}$ be so.

To prove the converse theorem that, if $\sum_{m=1}^{n_r} D_{r,m}$ be less than some fixed finite number, so also is $\sum_{m=1}^{n_r} \Delta_{r,m}$, let us first suppose that the function $f(x)$ has no external saltus at any point in the interval; it will then be proved that $\sum_{m=1}^{n_r} D_{r,m}$ and $\sum_{m=1}^{n_r} \Delta_{r,m}$, for a particular convergent system of sub-intervals, converge to the same limit, as r is indefinitely increased.

Consider the interval $(x_{r,m-1}, x_{r,m})$; let U and V be the upper and lower limits of $f(x)$ in this interval; thus $\Delta_{r,m} = U - V$. We have $D_{r,m} \leqq \Delta_{r,m}$; and it will be shewn that some greater value of r, say $r+s$, can be chosen, such that $\Sigma D_{r+s} \geqq \Delta_{r,m}$, where ΣD_{r+s} is the sum of the absolute values of the differences of the functional values at the end-points of the parts into which the interval $(x_{r,m-1}, x_{r,m})$ is divided, when the $(r+s)$th stage of the successive sub-divisions of (a, b) is reached. If U, V be the functional values at the ends of the interval $(x_{r,m-1}, x_{r,m})$ it is clear that $D_{r,m} = \Delta_{r,m}$; we therefore need only consider the case in which one at least of the numbers $f(x_{r,m-1}), f(x_{r,m})$ lies between U and V. If ϵ be an arbitrarily small positive number, there are two points ξ_1, ξ_2, in the interval, for one of which the functional value is greater than $U - \epsilon$, and for the other the functional value is less than $V+\epsilon$. Now, provided the function have no external saltus at any point, s may be chosen so large, and consequently d_{r+s} so small, that there is an end-point of one of the sub-intervals into which $(x_{r,m-1}, x_{r,m})$ is divided at the $(r+s)$th stage, such that at this end-point the functional value is $> U-2\epsilon$, and also such that there is another end-point at which the functional value is $< V+2\epsilon$: and one at least of these end-points does not coincide either with $x_{r,m-1}$ or with $x_{r,m}$. The number s having been so chosen, we see that for the interval $(x_{r,m-1}, x_{r,m})$, $\Sigma D_{r+s} > U-V-4\epsilon+W$, where W is the absolute difference between the functional value at one of the end-points of $(x_{r,m-1}, x_{r,m})$ and the functional value at one of those end-points of a sub-interval at which it is $> U-2\epsilon$, or else $< V+2\epsilon$. Since ϵ is arbitrarily small, it is clear that s can be chosen so great that $\Sigma D_{r+s} \geqq U-V$, for the interval in question. Moreover, s can be chosen so great that this condition is satisfied for each of the n_r intervals $(x_{r,m-1}, x_{r,m})$; and therefore s can be chosen such that ΣD_{r+s} taken for the whole interval (a, b) is $\geqq \sum_{m=1}^{n_r} \Delta_{r,m}$. It follows that the limit, as r is indefinitely increased, of $\sum_{m=1}^{n_r} D_{r,m}$ is $\geqq$ the limit of $\sum_{m=1}^{n_r} \Delta_{r,m}$; hence, since the first of these limits is $\leqq$ the second, it is seen that the two limits must be identical.

It has now been established that, *for a function with limited total fluctuation, and without points at which there is an external saltus, the total fluctuation and the total variation of the function are identical, being independent of any particular convergent system of sub-divisions.*

Next let $f(x)$ have an external saltus at each point of some set. If we consider a convergent system of sub-divisions such that no point of this set is ever an end-point of a sub-interval, it is clear that the total variation of $f(x)$ for such a system is identical with the total fluctuation of that function $\phi(x)$, which differs from $f(x)$ only in having the functional values at the points of the set so altered that the external saltus is at every point removed. It has been shewn that the total fluctuation of $\phi(x)$ is $2\mu-M$. If, on the other hand, a convergent system of sub-divisions be chosen, so that every point at which there is an external saltus becomes, at some stage, an end-point of a sub-interval, the total variation will be identical with the total fluctuation, their common value being M. It thus appears that, *for a function of limited total fluctuation, which has points with an external saltus, the total variation is M or $2\mu-M$, or has some value between these two numbers, according to the particular system of sub-divisions employed.*

The necessary and sufficient conditions that a function $f(x)$ defined for the interval (a, b) may be a function with limited total fluctuation, may be now stated as follows:—

(1) The points of discontinuity must all be of the first species, *i.e.* $f(x+0)$, $f(x-0)$ must everywhere exist.

(2) The sum of the absolute values of the external saltuses must be finite.

(3) A convergent system of sub-intervals must exist such that

$$\Sigma D_{r,m} = \Sigma \, | f(x_{r,m-1}) - f(x_{r,m}) |$$

is, for every value of r, less than some fixed number.

These conditions are clearly equivalent to those which have been given in the definition of the class of functions with limited total variation.

EXAMPLES.

1.* The function defined by $f(x) = x \sin \frac{1}{x}$, $f(0)=0$, is not of limited total fluctuation in the interval $(0, 1/\pi)$, although it is continuous in the interval. For in the interval $\left(\frac{1}{\overline{r+1}\,\pi}, \frac{1}{r\pi}\right)$, $\sin \frac{1}{x}$ attains the value $(-1)^r$, and thus the fluctuation in this interval is at least equal to $1/(r+\frac{1}{2})\pi$. The total fluctuation in the interval $\left(\frac{1}{s\pi}, \frac{1}{\pi}\right)$ is at least $\frac{1}{\pi}\left\{\frac{1}{1+\frac{1}{2}} + \frac{1}{2+\frac{1}{2}} + \dots + \frac{1}{s-1+\frac{1}{2}}\right\}$ or $\frac{2}{\pi}\left(\frac{1}{3} + \frac{1}{5} + \dots + \frac{1}{2s-1}\right)$, and it is well known that this increases without limit when s is indefinitely increased; therefore the total fluctuation in $(0, 1/\pi)$ is not finite.

2.† The function defined by $f(x) = x^2 \sin \frac{1}{x^2}$, $f(0)=0$, is continuous in any interval containing $x=0$, and is everywhere differentiable, but is not of limited total fluctuation.

3.* The function defined by $f(x) = x^2 \sin (x^{-\frac{4}{3}})$, $f(0)=0$, is of limited total fluctuation in the interval $(0, 1/\pi^{\frac{3}{4}})$. In the interval $\left(\frac{1}{(\overline{r+1}\,\pi)^{\frac{3}{4}}}, \frac{1}{(r\pi)^{\frac{3}{4}}}\right)$, the function has a single maximum, or else a single minimum, and the absolute value of the function at this point is at most $1/(r\pi)^{\frac{3}{2}}$. The total fluctuation in $(0, 1/\pi^{\frac{3}{4}})$ cannot exceed $2\sum_1^\infty \frac{1}{(r\pi)^{\frac{3}{2}}}$, which is finite.

4. Every function defined for a finite interval, which is continuous and of limited total fluctuation in that interval, is the difference of two continuous functions each of which is monotone in the interval.

THE MAXIMA, MINIMA, AND LINES OF INVARIABILITY OF CONTINUOUS FUNCTIONS.

197. Consider a point x_1 within the interval (a, b), in which a continuous function is defined; it may happen that a neighbourhood $(x_1 - \delta, x_1 + \delta)$ of the point x_1 can be found by taking δ sufficiently small, which is such that $f(x)$

* Lebesgue, *Leçons sur l'intégration*, p. 56.

† Lebesgue, *Annali di Mat.*, Ser. III A, vol. VII, p. 270.

has the same value at all points in the neighbourhood; then the point x_1 is called a *point of linear invariability* of the function. If the same holds for a neighbourhood of x_1 on the right only, or on the left only, then the point x_1 is called a *limiting point of linear invariability.*

It can be shewn that if a point x_1 of linear invariability exist, and the function be not constant in the whole interval (a, b), then there exist two limiting points of linear invariability, one of which, however, may be at one of the ends of the interval (a, b). Suppose the function not to be constant throughout the interval (x_1, b); the points x of this interval may be divided into two classes, in one of which x is such that in the interval (x_1, x) the function has the constant value $f(x_1)$, and in the other class x is such that (x_1, x) contains points at which the function has values differing from $f(x_1)$; a section is thus made of the interval (x_1, b), that defines a point which is the required limiting point of the linear invariability. If the same argument be applied to the interval (a, x_1) we see that there is another limiting point in this interval, unless the function be throughout equal to $f(x_1)$.

In the interval (a, b) there may be a finite number, or an indefinitely great, but enumerable, set of lines of invariability; each point within such a line is a point of invariability, and the ends of such lines are limiting points of invariability.

If the point x_1 be not a point of invariability, it may happen that a neighbourhood $(x_1 - \epsilon, x_1 + \epsilon')$ exists such that, for every point in the interior of this neighbourhood not identical with x_1, the condition $f(x) < f(x_1)$ is satisfied; in this case x_1 is said to be a point at which the function has a *proper maximum.* In case the neighbourhood be such that at every point x within it, except at x_1, the condition $f(x) > f(x_1)$ is satisfied, the point x_1 is said to be a point at which the function has a *proper minimum.*

It may happen that when x_1 is not a proper maximum, a neighbourhood $(x_1 - \epsilon, x_1 + \epsilon')$ exists which is such that at no point within it the condition $f(x) > f(x_1)$ is satisfied, nor at every point the condition $f(x) < f(x_1)$ is satisfied; in this case x_1 is said to be a point at which there is an *improper maximum* of the function. If the condition $f(x) \geqq f(x_1)$ is satisfied, but the condition $f(x) > f(x_1)$ is not everywhere satisfied, then x_1 is said to be a point at which there is an *improper minimum* of the function.

A line of invariability of which the end-points are α, β, and are both interior to (a, b), is said to be a maximum of the function, if both α, β be improper maxima, and it is said to be a minimum, if both α, β be improper minima.

It is clear that, in any arbitrarily small neighbourhood of an improper maximum or minimum, there are an indefinitely great number of points at which the functional value is equal to that at the maximum or minimum.

At any maximum or minimum there is a greatest neighbourhood $(x_1 - \delta, x_1 - \delta')$ at every interior point of which the condition $f(x) < f(x_1)$, $f(x) \leqq f(x_1)$, or $f(x) > f(x_1)$, $f(x) \geqq f(x_1)$ is satisfied. At end-points of such greatest neighbourhood, it follows from the condition of continuity of the function, that the functional value is equal to $f(x_1)$, unless the end-point coincides with a or with b.

It has been shewn in § 171 that there exists either one point or a set of points in (a, b) such that the functional value at this point or at all the points of the set is greater than at all other points in the interval; and it is to be remarked that this set of points may contain lines of invariability. Every such point, unless it be an end-point, is said to be a point of *absolute maximum* of the function in the interval (a, b), and may be either a proper or an improper maximum. A similar definition applies to an *absolute minimum.*

In case an extreme point of the continuous function (see § 167) be at a, or at b, such point is spoken of as an upper or lower extreme, but not always as a maximum or minimum of the function. If $f(a)$ and $f(b)$ be equal, and the function be not constant in (a, b), then there is at least one maximum or one minimum point, or one line of invariability, in the interior of (a, b). This is also true when $f(a) \neq f(b)$, unless the function be monotone.

198. *If within the interval (a, b) there be two points or two lines of invariability at which the function is a maximum, proper or improper, then there is between them at least one point or one line of invariability at which the function is a proper or improper minimum; thus maxima and minima occur alternately.*

Suppose that α, β are two points at which the function is a maximum, and that (α, β) is not entirely a line of invariability, also that no maximum occurs between α and β. We know that between α and β there is a point or a set of points at which the function is less than at all other points in the sub-interval; and since α and β cannot belong to such set, there is therefore a minimum at a point, or at points on a line of invariability, between α and β, and this minimum is less than either of the maxima at α and β.

Between a maximum and the next minimum of a function the function is said to make an oscillation, the amplitude of which is the excess of the maximum over the minimum.

If x_1 be a point in (a, b), it may be possible to choose ϵ so small that within the interval $(x_1, x_1 + \epsilon)$ no maxima or minima occur, so that the function is monotone in this interval. It may however be the case that, however small ϵ is taken, there still occur maxima and minima in $(x_1, x_1 + \epsilon)$. In this case the number of oscillations of the function must be indefinitely great, however small ϵ may be chosen; for if there were a finite number only, a number ϵ_1 could be found such that all the maxima and minima were in the interval $(x_1+\epsilon_1, x_1+\epsilon)$, and thus in $(x_1, x_1+\epsilon_1)$ the function would be monotone, which is contrary to the hypothesis made.

It thus appears that, in the neighbourhood of a particular point, a continuous function may have an indefinitely great number of oscillations. An improper maximum or minimum, not in a line of invariability, is certainly such a point.

The proper maxima and minima of a continuous function form an enumerable, or a finite, set of points.

Consider $(x_1-\epsilon, x_1+\eta)$, the greatest neighbourhood of a point of proper maximum x_1, which is such that for all other points x within the neighbourhood, $f(x) < f(x_1)$. There can in a finite interval be only a finite number of such points x_1 for which $\epsilon > \alpha$, $\eta > \alpha$, where α is a fixed positive number; for if there were an infinite number of such points, they would have a limiting point ξ, and we could choose two points x_1', x_1'' of the set, such that the distance of each from ξ is less than $\frac{1}{2}\alpha$; now each of these points would lie within the neighbourhood belonging to the other, and thus we should have $f(x_1') > f(x_1'')$, and also $f(x_1'') > f(x_1')$, which is impossible; thus the set must be finite. Now choose a sequence of descending values of α which converges to zero, say $\alpha_1, \alpha_2, \ldots \alpha_n, \ldots$; the number m_n of maxima x_1 such that for each $\epsilon > \alpha_n$, $\eta > \alpha_n$ being finite, we have $m_1, m_2, \ldots m_n, \ldots$ all finite; and hence the whole set of maxima forms an enumerable set.

If x_1 be an improper maximum point, and $f(x_1) = A$, a neighbourhood $(x_1-\epsilon, x_1+\eta)$ can be found which contains an infinite set of points G_A such that $f(x) = A$, for each point of the set. If x' be an isolated point of the set G_A, then x' is clearly a proper maximum of the function; and if x'' be a point of G_A, which is a limiting point of the set, x'' is an improper maximum. The points $x_1-\epsilon$, $x_1+\eta$ need not be maxima, even though they be limiting points of G_A. The condition of continuity of the function ensures that the set G_A is a closed one; for, at any limiting point of the set, the functional value is the limit of a sequence, each member of which is A, and this value is therefore itself A.

Corresponding to a given A, there may be a finite, or an infinite, set of detached intervals such as $(x_1-\epsilon, x_1+\eta)$, each one of which contains a closed set such that each isolated point of it is a proper maximum, and each limiting point (except an end-point) is an improper maximum. The sets G_A may contain perfect components, and thus the improper maxima at which A is the functional value may form a set of the cardinal number of the continuum. A similar result holds for minima.

It can further be shewn that the values of a continuous function at all its maxima and minima form a set which is either finite or enumerably infinite.

199. If in the interval (a, b) the function have only a finite number of maxima and minima, counting any line of invariability which is a maximum or minimum as one maximum or minimum, the interval can be divided into a finite number of parts in each of which the function is monotone; the function is then said to be* *in general monotone* (abtheilungsweise monoton).

If the function have an indefinitely great number of maxima and minima, which occur either at points or at lines of invariability, the function then makes an infinite number of oscillations; and these may occur in the neighbourhoods either of a finite number of points, or of an infinite number of points.

It can be shewn that in the case of a continuous function, although there may be an infinite number of oscillations of the function, there can be only a finite number of which the amplitude exceeds an arbitrarily small fixed number σ.

For it has been shewn in § 175 that a number ϵ can be determined, such that, in any sub-interval of length ϵ, the fluctuation of the function does not exceed σ; therefore in each of the sub-intervals

$$(a, a+\epsilon), \quad (a+\epsilon, a+2\epsilon), \ldots (a+n\epsilon, b),$$

the fluctuation of the function is not greater than σ. It follows that no oscillation of the function which is greater than σ can be completed in one of these sub-intervals, and that such an oscillation must require two at least of these sub-intervals for its completion; hence the number of such oscillations in (a, b) cannot exceed the finite number n. As the number σ is diminished indefinitely, it may happen that the number of oscillations of which the amplitude exceeds σ is increased indefinitely.

THE DERIVATIVES OF FUNCTIONS.

200. If a function $f(x)$ be defined for all points in the interval (a, b), then for a point x_1 in this interval we may regard the function $\dfrac{f(x)-f(x_1)}{x-x_1}$ as a function $F(x)$ of x, which is defined for all values of x in (a, b), except for the point x_1. This function $F(x)$, although undefined at the point x_1, has finite or infinite functional limits at that point, in accordance with the definitions in § 176.

If the limits $F(x_1+0)$, $F(x_1-0)$ both exist and have the same finite value, this value is called the *differential coefficient* at x_1 of the function $f(x)$. At the point a, if $F(a+0)$ exists, it is frequently said to be the differential coefficient of $f(x)$ at a; and at the point b, if $F(b-0)$ exists, it is said to be the differential coefficient of $f(x)$ at b.

* This term is due to C. Neumann; see his work *Ueber die nach Kreis- Kugel- und Cylinderfunctionen fortschreitenden Reihen.*

The condition that $f(x)$ may possess a differential coefficient at x_1 is that, corresponding to each arbitrarily chosen positive number ϵ, a neighbourhood $(x_1-\delta,\ x_1+\delta)$ can be found, such that

$$\left|\frac{f(\xi)-f(x_1)}{\xi-x_1}-\frac{f(\xi')-f(x_1)}{\xi'-x_1}\right|<\epsilon$$

for every pair of points ξ, ξ' which lie within this neighbourhood, or within such part of it as is interior to (a, b).

In other words, the condition is that a neighbourhood of x_1 can be found such that the fluctuation of the function $\frac{f(x)-f(x_1)}{x-x_1}$ within it, or within such part of it as lies in (a, b), may be as small as we please.

When a differential coefficient of $f(x)$ exists at the point x_1, then the function is said to be *differentiable* at x_1, and the differential coefficient at that point may be denoted by $f'(x_1)$.

That a function $f(x)$ may be differentiable at x_1, it is necessary, but not sufficient, that x_1 should be a point of continuity of the function.

At a point of discontinuity x_1 of $f(x)$, there always exists a positive number σ, such that in any neighbourhood of x_1, however small, points ξ exist such that $|f(\xi)-f(x_1)|>\sigma$; hence if A be any arbitrarily great positive number, in the interval $(x_1-\delta,\ x_1+\delta)$, where $\delta<\frac{\sigma}{A}$, there exist points ξ such that $\left|\frac{f(\xi)-f(x_1)}{\xi-x_1}\right|>A$, and it is thus impossible that $\frac{f(x)-f(x_1)}{x-x_1}$ should have a definite finite limit at x_1. On the other hand, the condition of differentiability, viz. that $\frac{f(x)-f(x_1)}{x-x_1}$ should have an arbitrarily small fluctuation within a sufficiently small neighbourhood of x_1, is not necessarily satisfied when the condition of continuity, viz. that $f(x)$ should have an arbitrarily small fluctuation within a sufficiently small neighbourhood of x_1, is satisfied.

It may happen that the limit of $\frac{f(x)-f(x_1)}{x-x_1}$, on both sides of x_1, is indefinitely great with the same sign on the two sides; in this case it is usual to say that $f(x)$ has a differential coefficient at x_1 which is infinite in value.

A continuous function $f(x)$ defined for the interval (a, b), which has a differential coefficient at every point of the interval, is said to be *differentiable in its domain*. Continuous functions exist, which at no point in their domain possess a differential coefficient. The first example of such a function was given by Weierstrass; the construction of such functions will be considered in Chapter VI.

That a continuous function possesses a differential coefficient was formerly regarded as obvious from geometrical intuition, it being supposed that such functions were necessarily representable by curves possessing definite tangents at every point. The first attempt to prove the existence of a differential coefficient of a continuous function was that of Ampère*; this proof was, however, insufficient even in the case of those continuous functions which make only a finite number of oscillations in the intervals for which they are defined. It is now fully recognized that the class of continuous functions is much wider than that of functions capable of an approximate graphical representation; and that the conditions for the existence of definite differential coefficients are of a much more stringent character than would be the case if they were included under the bare condition of continuity of the function.

201. It may happen that at a point x_1, the function $\dfrac{f(x)-f(x_1)}{x-x_1}$ possesses finite, or even indefinitely great, limits on the right and on the left at x_1 which differ from one another; the function is then said to have *derivatives, on the right* and *on the left* at x_1. These are frequently spoken of as the *progressive* and *regressive derivatives* respectively. A function may possess a progressive derivative and no regressive derivative, or the reverse.

When at the point x_1 a function is not differentiable, and possesses neither a derivative on the right nor one on the left, then the function $\dfrac{f(x)-f(x_1)}{x-x_1}$ has at x_1 four functional limits, an upper and a lower on the right, and an upper and a lower on the left; and any one of these may be either finite or infinite. These four limits are defined to be the *upper and lower derivatives* at x_1 *on the right*, and *the upper and lower derivatives* at x_1 *on the left*, and are, in accordance with the notation of Scheeffer†, denoted by $D^+f(x_1)$, $D_+f(x_1)$, $D^-f(x_1)$, $D_-f(x_1)$ respectively.

It is frequently convenient in this general case to speak of the derivatives of $f(x)$ on the right and on the left as existent but indefinite in value: and in this case $D^+f(x_1)$, $D_+f(x_1)$ are regarded as the limits of indeterminacy of the derivative on the right, and $D^-f(x_1)$, $D_-f(x_1)$ as those of the derivative on the left.

The definitions which have been given for the case in which the domain of the function is continuous are applicable, without essential change, to the case in which the domain is any perfect set of points. At a point of the set which is a limiting point on both sides there exist in general the four derivatives $D^+f(x)$, $D_+f(x)$, $D^-f(x)$, $D_-f(x)$, two or more of which may have

* *Journ. École polyt.*, vol. VI, 1806, p. 148.

† *Acta Mathematica*, vol. V. The same limits were considered by Du Bois Reymond, *Programm*, Freiburg, 1870, also *Münch. Abh.* vol. XII, p. 125, under the name Unbestimmtheitsgrenzen.

equal values; and at a point of the perfect set, which is a limiting point on one side only, there exist of course only the two derivatives on that side. If the domain be any closed set, the derivatives exist only at those points which are limiting points of the set.

A function defined for a perfect set may, by the method of correspondence, be correlated with a function defined for a continuous interval, the order of the points in the continuous interval and in the perfect set being the same; and thus all properties of derivatives of functions defined for a continuous interval have their analogues in the case in which the domain is any perfect set.

EXAMPLES.

1. If $f(x)=x\sin\frac{1}{x}$, $f(0)=0$; we have $\frac{f(h)-f(0)}{h}=\sin\frac{1}{h}$, and for arbitrarily small values of h, this oscillates between 1 and -1. The function $f(x)$, although continuous at $x=0$, possesses no differential coefficient at that point; in fact

$$D^+f(0)=1,\ \ D_+f(0)=-1,\ \ D^-f(0)=1,\ \ D_-f(0)=-1.$$

2. If $f(x)=x^2\sin\frac{1}{x}$, $f(0)=0$, the differential coefficient $f'(x)$ exists for every value of x, and is finite. At the point $x=0$, $f'(x)$ is zero, but has a discontinuity of the second kind.

3. Let* $f(x)=\sqrt{x}\left(1+x\sin\frac{1}{x}\right)$, for $x>0$; $f(x)=-\sqrt{-x}\left(1+x\sin\frac{1}{x}\right)$, for $x<0$; and $f(0)=0$. In this case $f'(x)$ everywhere exists; its value at $x=0$, is $+\infty$, and although it has a finite value at every point except at $x=0$, it oscillates in the neighbourhood of that point between indefinitely great positive and negative values.

4.† The function defined by $f(x)=x\{1+\frac{1}{3}\sin(\log x^2)\}$, and $f(0)=0$, is everywhere continuous, and is monotone, but has no differential coefficient at $x=0$.

5.‡ Let $f(x)=e^{-\frac{1}{x^2}}\sin\frac{1}{x}$, $f(0)=0$; this function has at every point a differential coefficient, and this is continuous at $x=0$. The differential coefficient vanishes at $x=0$, and at an infinite number of points in the neighbourhood of $x=0$. The function $f'(x)$ like $f(x)$, has an infinite number of oscillations in a neighbourhood of $x=0$.

THE DIFFERENTIAL COEFFICIENTS OF CONTINUOUS FUNCTIONS.

202. Let us suppose that a continuous function, defined for a continuous domain, is such that at every point interior to an interval (α, β) there exists a differential coefficient; this differential coefficient may at any point have a finite value which may be zero, or it may have an infinite value of which, however, the sign is definite. It will be observed that $f(x)$ is assumed to be

* Dini, *Grundlagen*, p. 112.

† Pringsheim, *Encyklopädie der Math. Wissensch.* II A. I, p. 22.

‡ Dini, *Grundlagen*, p. 313.

continuous at the points α, β, but it is not assumed that definite derivatives exist at those points. It will be shewn that, *unless the function be constant throughout* (α, β), *there exists at least one point in the interior of* (α, β) *at which the differential coefficient has a definite finite value different from zero.*

Suppose $f(\alpha)$, $f(\beta)$ to be unequal. If they be not unequal, and the function be not constant throughout (α, β), we can replace the interval (α, β), by another one contained in it, for which the functional values at the ends are unequal. Let us consider the function

$$F(x) = f(x) - f(\alpha) - \frac{x-\alpha}{\beta-\alpha}\{f(\beta) - f(\alpha)\}.$$

$F(\alpha)$ and $F(\beta)$ vanish, and $F(x)$ is continuous in (α, β), and has a differential coefficient in the ordinary sense at each point, with the possible exception of α and β; therefore it follows by the theorem of § 171 that there is at least one point x_1 in the interior of (α, β), at which $F(x)$ is a maximum or minimum: this is the case even if $F(x)$ be everywhere zero in the interval. A number ϵ can therefore be found such that $F(x_1+\delta) - F(x_1)$, $F(x_1-\delta) - F(x_1)$ have the same sign, or else vanish, provided $\delta < \epsilon$; and consequently the derivatives at x_1 on the right and left must have opposite signs, unless both of them be zero; therefore the differential coefficient at x_1, which must exist, must be zero. It follows that $f'(x_1) - \dfrac{f(\beta) - f(\alpha)}{\beta-\alpha} = 0$, and thus the point x_1 is the point of which the existence was to be proved. From this theorem we deduce the following general theorem:—

If $f(x)$ *be continuous in the interval* (a, b), *and be such that it has a differential coefficient at every point in the interior of the interval, and if there be in* (a, b) *no lines of invariability, then there exists in* (a, b) *an everywhere-dense set of points at which the differential coefficient has finite values differing from zero.*

This is proved at once by applying the foregoing theorem to any interval contained in (a, b). There may be in (a, b) infinite sets of points at which the differential coefficient is either zero or infinite.

203. *If the function* $f(x)$ *be continuous in the interval* $(x, x+h)$, *and at every point in the interior of this interval* $f'(x)$ *exist, being either finite or infinite with fixed sign, then a point* $x+\theta h$ *exists, where* θ *is some proper fraction, and is neither* 0 *nor* 1, *such that*

$$f(x+h) = f(x) + hf'(x+\theta h).$$

This is at once seen by taking $\alpha = x$, $\beta = x+h$, $x_1 = x+\theta h$ in the proof in § 202. This is known as *the mean value theorem* of the Differential Calculus.

A corollary from the mean value theorem is that, if $f(x) = f(x+h)$, then $f'(x)$ must be zero at one point at least in the interior of the interval

$$(x,\ x+h).$$

An important extension of the mean value theorem is the following:—

If $f(x)$ be continuous in the interval $(x, x+h)$, and have a differential coefficient at every point of the interval, with the possible exception of the end-points; and if $F(x)$ be another function which is also continuous in the same interval, and at every interior point has a finite differential coefficient different from zero, whilst at the end-points there may be no definite derivatives, or they may be zero, or infinite, then

$$\frac{f(x+h)-f(x)}{F(x+h)-F(x)}=\frac{f'(x+\theta h)}{F'(x+\theta h)}$$

for some value of θ which is a proper fraction, and is neither 1 *nor* 0.

To prove the theorem, let

$$\phi(\xi)=f(\xi)-f(\alpha)-\frac{f(x+h)-f(x)}{F(x+h)-F(x)}\{F(\xi)-F(x)\};$$

then, since $F'(x)$ does not vanish in the interior of the interval $(x, x+h)$, it follows that $F(x+h)-F(x)$ cannot be zero. Since $\phi(x)=\phi(x+h)$, and $\phi(\xi)$ satisfies the conditions of the mean value theorem, $\phi'(\xi)$ must vanish for some value $x+\theta h$ of ξ, interior to the interval $(x, x+h)$. We have then

$$f'(x+\theta h)-\frac{f(x+h)-f(x)}{F(x+h)-F(x)}F'(x+\theta h)=0,$$

from which the theorem follows, since $F'(x+\theta h)$, and therefore $f'(x+\theta h)$ cannot be infinite. In the case in which $f(x+h)=f(x)$, we have

$$f'(x+\theta h)=0,$$

for some suitable value of θ; and then, since $F(x+h)-F(x)$, $F'(x+\theta h)$ are finite, the theorem still holds.

204. The last theorem may be applied to obtain a strict proof of the legitimacy, under certain conditions, of a well-known method of evaluating limits which appear in the so-called indeterminate forms $\frac{0}{0}, \frac{\infty}{\infty}$.

Let the two functions $f(x)$, $F(x)$ be both continuous at all points interior to the interval $(\alpha, \alpha+\beta)$, and let the limits $f(\alpha+0)$, $F(\alpha+0)$ both exist and be zero; if finite differential coefficients $f'(x)$, $F'(x)$ exist at every interior point of $(\alpha, \alpha+\beta)$, and $F'(x)$ be everywhere within this interval different from zero, then if one of the two limits

$$\lim_{h=0}\frac{f(\alpha+h)}{F(\alpha+h)},\quad \lim_{h=0}\frac{f'(\alpha+h)}{F'(\alpha+h)}$$

exist as a definite number, or be infinite with a fixed sign, the other limit also exists, and the two have the same value.

The two functional values $f(\alpha)$, $F(\alpha)$ may both be defined to be zero, and thus the functions $f(x)$, $F(x)$ are continuous in any interval $(\alpha, \alpha+h)$, when $h<\beta$. We have then, from the extension of the mean value theorem

$$\frac{f(\alpha+h)}{F(\alpha+h)}=\frac{f'(\alpha+\theta h)}{F'(\alpha+\theta h)},$$

where θ is some proper fraction. Since θh converges to zero when h does so, the theorem follows at once from this equality.

Let the two functions $f(x)$, $F(x)$ be both continuous at all points interior to the interval $(\alpha, \alpha+\beta)$, and let the limits $f(\alpha+0)$, $F(\alpha+0)$ both exist and be infinite, each with a fixed sign; if finite differential coefficients $f'(x)$, $F'(x)$ exist at every interior point of $(\alpha, \alpha+\beta)$, and $F'(x)$ be everywhere, within this interval, different from zero, then if one of the two limits*

$$\lim_{h=0}\frac{f(\alpha+h)}{F(\alpha+h)},\quad \lim_{h=0}\frac{f'(\alpha+h)}{F'(\alpha+h)}$$

exist as a definite number, or be infinite with a fixed sign, the other limit also exists, and the two have the same value.

Consider the interval $(\alpha+\delta_1,\ \alpha+\delta_2)$ interior to $(\alpha, \alpha+\beta)$; we have then

$$\frac{f(\alpha+\delta_2)-f(\alpha+\delta_1)}{F(\alpha+\delta_2)-F(\alpha+\delta_1)}=\frac{f'(\alpha+\delta_3)}{F'(\alpha+\delta_3)},$$

where δ_3 lies between the numbers δ_1 and δ_2: this equation may be written in the form

$$\frac{f(\alpha+\delta_1)}{F(\alpha+\delta_1)}=\frac{f(\alpha+\delta_2)}{F(\alpha+\delta_1)}+\frac{f'(\alpha+\delta_3)}{F'(\alpha+\delta_3)}\left\{1-\frac{F(\alpha+\delta_2)}{F(\alpha+\delta_1)}\right\}.$$

Taking a fixed value of δ_2, and an arbitrarily small positive number ϵ, we can find a positive number $\delta'\ (<\delta_2)$, such that for every value of $\delta_1\ (>0)$ which is $<\delta'$, the inequalities

$$|F(\alpha+\delta_1)|>\frac{1}{\epsilon}|f(\alpha+\delta_2)|,\quad |F(\alpha+\delta_1)|>\frac{1}{\epsilon}|F(\alpha+\delta_2)|$$

are both satisfied: this follows from the fact that $F(\alpha+0)$ is infinite with fixed sign.

We have now

$$\frac{f(\alpha+\delta_1)}{F(\alpha+\delta_1)}=\eta+(1-\zeta)\frac{f'(\alpha+\delta_3)}{F'(\alpha+\delta_3)},$$

where $|\eta|<\epsilon$, and $|\zeta|<\epsilon$, for all values of δ_1 which are >0, and $<\delta'$.

Let us first assume that $\dfrac{f'(\alpha+h)}{F'(\alpha+h)}$ has a definite finite limit k; we may

* The unnecessary hypothesis is made by Stolz (see *Grundzüge*, vol. I, p. 77), that $F'(x)$ has everywhere the same sign. For a history of these theorems, see Pringsheim, *Encyklopädie d. Math. Wissensch.* II A. I, p. 26.

then choose δ_2 so small that $\frac{f'(a+\delta_3)}{F'(a+\delta_3)}-k$ is numerically less than an arbitrarily chosen positive number η', for every value of δ_3, which is $< \delta_2$; we have then

$$\frac{f(a+\delta_1)}{F(a+\delta_1)}-k=\eta+\zeta'-\zeta(k+\zeta'), \text{ when } |\zeta'|<\eta'.$$

Since η' and ϵ are both arbitrarily small, the absolute value of

$$\frac{f(a+\delta_1)}{F(a+\delta_1)}-k$$

is arbitrarily small, for all sufficiently small values of δ_1; and it thus follows that

$$\lim_{h=0}\frac{f(a+\delta_1)}{F(a+\delta_1)}=k.$$

Next let us assume that $\frac{f'(a+h)}{F'(a+h)}$ has an infinite limit, for $h=0$; we may, without loss of generality, take this limit to be of positive sign. We may choose an arbitrarily large positive number N, and a number $N'>N$; and we may then choose δ_2 so small that $\frac{f'(a+\delta_3)}{F'(a+\delta_3)}>N'$, for all possible values of $\delta_3<\delta_2$; then

$$\frac{f(a+\delta_1)}{F(a+\delta_1)}=\eta+(1-\zeta)(N'+p),$$

where p is positive.

The number ϵ may now be chosen so small that

$$\eta+(1-\zeta)(N'+p)>N$$

for all the possible values of η and ζ; and therefore an interval on the right of a can be determined, such that for all interior points the inequality

$$\frac{f(a+h)}{F(a+h)}>N$$

is satisfied. Since N is arbitrarily great, it follows that

$$\lim_{h=0}\frac{f(a+h)}{F(a+h)}=+\infty.$$

The relation $$\frac{f'(a+\delta_3)}{F'(a+\delta_3)}=\frac{1}{1-\zeta}\frac{f(a+\delta_1)}{F(a+\delta_1)}-\frac{\eta}{1-\zeta}$$

may be employed to prove, in a similar manner, that the existence of

$$\lim_{h=0}\frac{f(a+h)}{F(a+h)}$$

involves that of $\lim_{h=0}\frac{f'(a+h)}{F'(a+h)}$, and that the two limits have the same value.

205. *If the otherwise continuous function $f(x)$ have a discontinuity of the second kind at the point a, at least on the side which is towards the interval $(a, a+h)$, but the function have a finite differential coefficient at every point of the interval $(a, a+h)$, except at the point a, then the absolute values of these differential coefficients in any arbitrarily small neighbourhood of a have no upper limit.*

By applying the mean value theorem we have

$$f(a+\delta_2)-f(a+\delta_1)=(\delta_2-\delta_1)f'(a+\delta_3),$$

where $0<\delta_1<\delta_2<h$, and δ_3 is some number lying between δ_1 and δ_2.

Now if $\overline{f(a+0)}$ and $\underline{f(a+0)}$ be unequal, values of δ_1 and δ_2, less than any arbitrarily prescribed positive number ϵ, can be chosen, such that

$$f(a+\delta_2)-f(a+\delta_1)$$

is arbitrarily near to $\overline{f(a+0)}-\underline{f(a+0)}$, whereas $\delta_2-\delta_1$ is arbitrarily small; therefore it follows that $f'(a+\delta_3)$ must have arbitrarily great values, in any neighbourhood of a.

The mean value theorem $f(a+h)-f(a)=hf'(a+\theta h)$, where $0<\theta<1$, affords information as to the existence and value of the derivative at a, on the right, provided $f(x)$ satisfies, in a neighbourhood of a on the right, the conditions under which the theorem holds. By considering both sides of a, information may be obtained as to the existence of a differential coefficient at a.

(1) If the function $f'(x)$ have a functional limit at a on the right, then

$$\frac{f(a+h)-f(a)}{h}$$

has a definite limit for $h=0$, either finite, or infinite with fixed sign, and this is equal to that of $f'(x)$. It follows that, in this case, a derivative at a on the right exists, and is either finite, or infinite with fixed sign.

(2) If the function $f'(x)$ have no limit at a on the right, it may still happen that $f'(a+\theta h)$ has a definite limit at a on the right, because $a+\theta h$ is not necessarily capable of having all values within a neighbourhood of a. In this case, either (a) the derivative at a on the right may be definite, and lie between the upper and lower limits of $f'(x)$ at a on the right, or it may be equal to one or other of those limits; or (b) there may be no definite derivative at a on the right, but $D^+f(a)$, $D_+f(a)$ may have different values, and these are certainly both finite in case the upper and lower functional limits of $f'(x)$ at a are both finite.

(3) The derivative on the right at a can only exist and be infinite, (a) if $f'(x)$ have an infinite limit on the right at a, or (b) if it have an infinite upper limit on the right at a. In either of the cases, (a) and (b), $f'(x)$ may be everywhere finite within a neighbourhood of a on the right, or it may be infinite at some points in such a neighbourhood.

(4) If the derivative at a on the right exist and be finite, then either (*a*) $f'(x)$ has a definite limit at a on the right, equal to the derivative at a, or (*b*) $f'(x)$ has no definite limit at a on the right, but a sequence of points can be determined, of which a is the limiting point, such that the values of $f'(x)$ for points of that sequence converge to the value of the derivative at a. At points which do not belong to the sequence, the values of $f'(x)$ may be either finite or infinite.

(5) The non-existence of a definite derivative at a on the right may be due to the non-existence of $f'(x)$ at all the points of any neighbourhood of a, or only at an infinite number of points of such a neighbourhood.

206. *If $f(x)$ be continuous in a given interval and have at every point, with the exception of an enumerable set G, a differential coefficient of value zero, the function is constant throughout the whole interval.*

At the points of G we may suppose it to be unknown whether a differential coefficient exists, or, if one does exist, what values it has.

A more general form of this theorem is obtained by considering not the differential coefficient, but any one of the four derivatives, thus:—

If $f(x)$ be continuous in a given interval, and one of the four derivatives $D^+f(x)$, $D_+f(x)$, $D^-f(x)$, $D_-f(x)$, be such that it is zero at every point of the interval, with the exception of points belonging to an enumerable set G, at which nothing is known as to its value, then the function is constant throughout the interval.

To prove the generalized theorem for the case of the function $D^+f(x)$, suppose that, if possible, $f(x)-f(a)$ has at some point x_1 a value different from zero, say the positive value p; and let $\phi(x, k)$ denote $f(x)-f(a)-k(x-a)$. Then $\phi(a, k)=0$, $\phi(x_1, k)=p-k(x_1-a)$. Choose any fixed positive number $q<p$, then $\phi(x_1, k)>q$, provided $k<\dfrac{p-q}{x_1-a}$, or say $k<K$. Since $\phi(x, k)$ is continuous in (a, b), and $\phi(a, k)$ is zero, whilst $\phi(x_1, k)>q$, there exists an upper limit of those values of x between 0 and x_1, for which $\phi(x, k)\leqq q$, and this upper limit is attained for some value ξ of x, which is such that $\xi<x_1$, and $\phi(\xi, k)=q$. Since $\phi(\xi+h, k)>q$, provided $0<h\leqq x_1-\xi$, we see that, since $\dfrac{\phi(\xi+h, k)-\phi(\xi, k)}{h}$ is positive, $D^+\phi(\xi, k)$ is positive if it be not zero. Now if ξ were a point not belonging to G, the value of $D^+\phi(\xi, h)$ would reduce to $-k$; and therefore ξ must belong to G.

The number q being fixed, ξ depends only on k; and, corresponding to a given value of ξ, there is only one value of k; for

$$\phi(\xi, k)-\phi(\xi, k')=(k'-k)(\xi-a),$$

which cannot vanish unless $k=k'$, since $\phi(a, k)$ is zero and therefore $<q$. For

a given value of k, the corresponding number of values of ξ, all of which necessarily belong to G, must be either finite or enumerably infinite, since every part of an enumerable aggregate is either finite or enumerable. Therefore to each value of k, in the continuous interval $(\alpha, K-\beta)$, there corresponds a finite or enumerable set of values of ξ, and it would hence follow that the continuum $(\alpha, K-\beta)$ is itself enumerable, which we know is not the case. It has thus been shewn that for no point can $f(x)-f(a)$ have a positive value; and similarly, by considering $f(x)-f(a)+k(x-a)$, it can be shewn that $f(x)-f(a)$ can nowhere have a negative value; hence $f(x)=f(a)$ throughout the whole interval (a, b). The case in which one of the other three derivatives vanishes except at points of G can be treated in a similar manner.

The following theorem* which is of importance in the theory of Integration will now be established:—

If two functions be each continuous in a given interval, and if of one of the four derivatives it be known that, for the two functions, this derivative has equal finite values at each point of the interval, with the exception of an enumerable set of points at which nothing is known as regards the two derivatives, then the two functions differ from one another only by a constant, which must be the same for the whole interval.

It must first be observed that the proof of the preceding theorem suffices to shew that, if $D^+f(x) \leqq 0$, at every point of (a, b) not belonging to the set G, then $f(x)-f(a) \leqq 0$, for every point x of the interval. Similarly, if $D^+f(x) \geqq 0$, everywhere in the interval, except at the points of G, then $f(x)-f(a) \geqq 0$ at every point of the interval.

If now $f_1(x)$, $f_2(x)$ be two continuous functions such that

$$D^+f_1(x)=D^+f_2(x)$$

at every point of (a, b) not belonging to G, let $f(x)=f(x_1)-f(x_2)$. If ϵ be an arbitrarily small positive number, then for any point x not belonging to G, the condition

$$\frac{f_1(x+h)-f_1(x)}{h} > D^+f_1(x)-\epsilon$$

is satisfied for a set of positive values of h which are arbitrarily small. Also we have, for all sufficiently small values of h,

$$\frac{f_2(x+h)-f_2(x)}{h} < D^+f_2(x)+\epsilon;$$

hence, since $D^+f_1(x)=D^+f_2(x)$, we see that $\dfrac{f(x+h)-f(x)}{h} > -2\epsilon$, for all

* Scheeffer, *Acta Mat.* vol. v, p. 283.

values of h belonging to some set. It follows that $D^+f(x) > -2\epsilon$, and thence* that $D^+f(x) \geqq 0$, since ϵ is arbitrary. By interchanging $f_1(x)$ and $f_2(x)$, we see that $D^+\{-f(x)\} \geqq 0$. From these two results we deduce that $f(x)-f(a) \geqq 0$, and that $f(a)-f(x) \geqq 0$, throughout the interval (a, b); therefore $f(x)$ is everywhere equal to $f(a)$, and thus the theorem is established.

207. At a point x at which the continuous function $f(x)$ is a maximum, since, for a sufficiently small neighbourhood of such a point x, the differences

$$f(x+h)-f(x), \quad f(x-h)-f(x)$$

are both negative or zero for all points $x \pm h$ in the neighbourhood, it is clear that each of the derivatives $D^+f(x)$, $D_+f(x)$ is either negative or zero, and that each of the derivatives $D^-f(x)$, $D_-f(x)$ is either positive or zero. In case the function possess definite derivatives on the right and on the left at the point x, the first of these is zero or negative, or possibly $-\infty$, whilst the second is zero or positive, or possibly $+\infty$.

If at the point x a definite differential coefficient exist, it must consequently be zero. In the case of a minimum the corresponding statements hold, where the positive sign takes the place of the negative one, and the reverse. The following theorem has now been established:—

If a continuous function possess a differential coefficient at a point x at which the function is a maximum or minimum, then the differential coefficient at x must be zero.

208. A continuous function may be such that in the interval (a, b) there exists an everywhere-dense set of non-overlapping intervals, each one of which is a line of invariability of the function. Within each interval of the set, the function has its differential coefficient equal to zero; it therefore follows from the theorem in § 206, that the closed set of points, of which the given set of intervals is the complementary set, cannot be an enumerable set, otherwise the function would be constant in the whole interval (a, b). It is further clear that no two of the intervals can abut on one another; for the condition of continuity of the function at their common end-point would ensure that the values of the function in the two intervals were the same, and thus the two intervals would really belong to the same line of invariability. It follows that the end-points and external points of an everywhere-dense set of lines of invariability of a continuous function must form a perfect non-dense set of points.

That a continuous function with an everywhere-dense set of lines of invariability can actually exist can be easily shewn as follows:—Make the points of a non-dense perfect set correspond in order to the points of

* It is erroneously stated by Dini, that $D^+f(x)=0$. See *Grundlagen*, p. 275.

a continuous interval (a, b), then, as has been shewn in § 128, the correspondence may be such that the whole of a complementary interval of the perfect set corresponds to one point of the continuous interval. If a continuous function be defined for the continuous interval, we may define a new function which has at each point of the perfect set the same value as the original function has at the corresponding point of the continuous interval; and since all the points of a complementary interval of the perfect set correspond to the same point of the continuous interval, the new function is such that it has an everywhere-dense set of lines of invariability.

EXAMPLES.

1. Take* the non-dense perfect set defined in Ex. 1, § 75, by

$$x=\frac{C_1}{3}+\frac{C_2}{3^2}+\ldots+\frac{C_n}{3^n}+\ldots,$$

where every C_n is either 0 or 2. A complementary interval has as its end-points

$$\frac{C_1}{3}+\frac{C_2}{3^2}+\ldots+\frac{C_{n-1}}{3^{n-1}}+\frac{1}{3^n},\quad \frac{C_1}{3}+\frac{C_2}{3^2}+\ldots+\frac{C_{n-1}}{3^{n-1}}+\frac{2}{3^n},$$

which may be denoted by (a_ν, b_ν). Let the function $f(x)$ be defined as follows:—For a point x of the interval (0, 1) belonging to the perfect set, let

$$f(x)=\frac{1}{2}\left(\frac{C_1}{2}+\frac{C_2}{2^2}+\ldots+\frac{C_n}{2^n}+\ldots\right);$$

when x is in the interval (a_ν, b_ν), let $f(x)=f(a_\nu)=f(b_\nu)$. The function $f(x)$ so defined is continuous, and varies from 0 to 1, and is constant in each of the intervals (a_ν, b_ν) complementary to the non-dense perfect set.

2.† Let the numbers in the interval (0, 1) be expressed in a scale $n\equiv 2m-1$, of odd degree; thus $x=\frac{a_1}{n}+\frac{a_2}{n^2}+\ldots$, where $0\leqq a_r<n$, and the number of digits a_r is finite or infinite. For any number x represented in this manner, for which all the a_r are even integers, let $f(x)$ equal $\frac{1}{2}\left(\frac{a_1}{m}+\frac{a_2}{m^2}+\ldots\right)$. In case any of the a_r are odd, let a_k be the first one which is odd, and let $f(x)$ then equal $\frac{1}{2}\left(\frac{a_1}{m}+\frac{a_2}{m^2}+\ldots+\frac{a_{k-1}}{m^{k-1}}\right)+\frac{1}{2}\frac{a_k+1}{m^k}$. This function $f(x)$ is continuous and varies from 0 to 1; for an infinite set of points it has no differential coefficient, and for all other values of x, $f'(x)=0$.

THE SUCCESSIVE DIFFERENTIAL COEFFICIENTS OF A CONTINUOUS FUNCTION.

209. If a continuous function $f(x)$, defined for the interval (a, b), have at every point a differential coefficient $f'(x)$, which is itself continuous throughout the interval, the function $f'(x)$ may itself have a differential coefficient $f''(x)$, which is called the second differential coefficient or derivative of $f(x)$.

* Cantor, *Acta Mat.* vol. IV, p. 386. See also Scheeffer, *Acta Mat.* vol. V, p. 289.

† Gravé, *Comptes Rendus*, vol. CXXVII, p. 1005.

The second differential coefficient of $f(x)$ at a point x_1, when it exists, is expressible as a repeated limit

$$\lim_{k=0}\left\{\lim_{h=0}\frac{f(x_1+h+k)-f(x_1+h)-f(x_1+k)+f(x_1)}{hk}\right\},$$

in which the limit for $h=0$ is to be first obtained, and then the limit for $k=0$.

The existence of the repeated limit as a definite number does not necessarily imply the existence of $f''(x_1)$.

The above definition of $f''(x)$ is applicable at any point x_1 for which a neighbourhood $(x_1-\epsilon, x_1+\epsilon')$ exists, such that $f'(x)$ exists everywhere in that neighbourhood and is continuous. The continuity of $f'(x)$ is however not sufficient to ensure that $f''(x)$ exists.

When $f'(x_1)$ has a definite value, but $f'(x)$ fails to possess a definite value at some or all of the points of the neighbourhood $(x_1-\epsilon, x_1+\epsilon')$, it may happen that the ratios

$$\frac{D^+f(x_1+k)-f'(x_1)}{k},\qquad \frac{D_+f(x_1+k)-f'(x_1)}{k},$$

$$\frac{D^-f(x_1+k)-f'(x_1)}{k},\qquad \frac{D_-f(x_1+k)-f'(x_1)}{k},$$

all have the same limit for $k=0$. In this case we may regard this limit as defining $f''(x_1)$; and thus this extended definition is applicable to cases in which $f'(x)$ exists at the point x_1, and at some only, or at none, of the points in any neighbourhood of x_1, however small that neighbourhood may be chosen.

210. *If in an interval, which contains in its interior the point x_1, the differential coefficient $f'(x)$ of $f(x)$ everywhere exist, and be continuous through the interval, and if further the second differential coefficient $f''(x)$ exist throughout the interval, being at every point either finite, or infinite with a definite sign, and be finite at the point x_1, then $f''(x_1)$ is the limit, when $h=0$, of either of the expressions*

$$\frac{f(x_1+h)-2f(x_1)+f(x_1-h)}{h^2},$$

$$\frac{f(x_1+2h)-2f(x_1+h)+f(x_1)}{h^2}.$$

The converse does not hold; for either of these expressions may have a definite finite limit at $h=0$, and yet $f''(x_1)$ may not exist, or even $f'(x_1)$

may not exist. An illustration of this is the case of the function defined by $f(0)=0$, $f(x)=x\sin^2\frac{1}{x}$ for $x^2>0$; at the point 0, $f'(0)$ has no existence, and yet

$$\lim_{h=0}\frac{f(h)-2f(0)+f(-h)}{h^2}=0.$$

To prove the theorem, we may take $(x_1-\epsilon, x_1+\epsilon)$ as the neighbourhood of the point x_1, through which $f'(x)$ is continuous and $f''(x)$ everywhere exists. Suppose $f''(x_1)=k$; and let us consider the function $\phi(x)=f(x)-\frac{1}{2}kx^2$, which has similar properties to the function $f(x)$; and thus $\phi''(x_1)=0$.

If $h<\epsilon$,

$$\phi(x_1+h)-\phi(x_1)=\quad h\phi'(x_1+\theta h),$$
$$\phi(x_1-h)-\phi(x_1)=-h\phi'(x_1-\theta_1 h),$$

where θ, θ_1 are proper fractions; again

$$\phi'(x_1+\theta h)\;-\phi'(x_1)=\quad \theta h\phi''(x_1+\theta\theta_2 h),$$
$$\phi'(x_1-\theta_1 h)-\phi'(x_1)=-\theta_1 h\phi''(x_1-\theta_1\theta_3 h),$$

where θ_2, θ_3 are proper fractions. We find from these results

$$\frac{\phi(x_1+h)-2\phi(x_1)+\phi(x_1-h)}{h^2}=\theta\phi''(x_1+\theta\theta_2 h)+\theta_1\phi''(x_1-\theta_1\theta_3 h).$$

Since $\phi''(x_1)$ exists and is zero,

$$\frac{\phi'(x_1+\theta h)-\phi'(x_1)}{\theta h},\qquad \frac{\phi'(x_1-\theta_1 h)-\phi'(x_1)}{-\theta_1 h},$$

both have zero as limit, for $h=0$; hence the same is true of $\phi''(x_1+\theta\theta_2 h)$, $\phi''(x_1-\theta_1\theta_3 h)$, as is seen from the formulae above.

It has thus been shewn that

$$\lim_{h=0}\frac{\phi(x_1+h)-2\phi(x_1)+\phi(x_1-h)}{h^2}=0,$$

which shews that

$$\lim_{h=0}\frac{f(x_1+h)-2f(x_1)+f(x_1-h)}{h^2}=k=f''(x_1).$$

A similar proof establishes the theorem which relates to the other limit.

211. The following theorem, due to Schwarz*, is of fundamental importance in the theory of Fourier's series.

If, in an interval (α, β) in which $f(x)$ is continuous, the expression

$$\frac{f(x+h)-2f(x)+f(x-h)}{h^2}$$

converge for each value of x in (α, β) to the limit zero, for $h=0$, then the function $f(x)$ is a linear function in the whole interval, and consequently $f'(x)$, $f''(x)$ everywhere exist, and the latter is everywhere zero.

* *Crelle's Journal*, vol. LXXII.

Let us consider the function

$$\phi(x) = \pm \left\{ f(x) - f(\alpha) - \frac{x-\alpha}{\beta-\alpha}[f(\beta)-f(\alpha)] \right\} + k^2 (x-\alpha)(x-\beta),$$

where k is a constant. The function $\phi(x)$, whichever sign be taken, is continuous in (α, β), and vanishes at α and β. We find at once

$$\lim_{h=0} \frac{\phi(x+h) - 2\phi(x) + \phi(x-h)}{h^2} = 2k^2;$$

and therefore, for each value of x in (α, β), a positive number ϵ can be found, such that $\phi(x+h) - 2\phi(x) + \phi(x-h)$ is positive and greater than zero for all values of h which are numerically less than ϵ.

If $\phi(x)$ could be anywhere positive in (α, β), there must be a point x_1 at which it has the greatest positive value, and this point is not α nor β, since $\phi(\alpha)$, $\phi(\beta)$ both vanish. If η be sufficiently small,

$$\phi(x_1+\eta) - \phi(x_1) \leqq 0, \quad \phi(x_1-\eta) - \phi(x_1) \leqq 0,$$

hence
$$\phi(x_1+\eta) - 2\phi(x_1) + \phi(x_1-\eta)$$

would be, for all sufficiently small values of η, either negative or zero, which is contrary to what was shewn above. It follows that $\phi(x)$ is everywhere negative in (α, β), and cannot be zero except at α and β.

This holds whichever sign be taken in defining $\phi(x)$. Now $k^2(x-\alpha)(x-\beta)$ is always negative except at α and β, and may be taken to have its numerically greatest value as small as we please, since k is at our choice. It follows that

$$f(x) - f(\alpha) - \frac{x-\alpha}{\beta-\alpha}[f(\beta)-f(\alpha)]$$

can nowhere in the interval be different from zero; for, if at any point it had a value p, by choosing k such that $k^2(x-\alpha)(x-\beta)$ is numerically everywhere $< p$, the function $\phi(x)$ could be made positive at the point by proper choice of the ambiguous sign. It has thus been shewn that $f(x)$ is linear in (α, β).

212. Schwarz's theorem can be extended to the case in which there is an enumerable set of points in the interval (α, β), at which it is not known that the limit in question exists, or is zero, provided a certain condition be satisfied at each point of the enumerable set. The following theorem will be established:—

If, in an interval (α, β) in which $f(x)$ is continuous, the expression

$$\frac{f(x+h) - 2f(x) + f(x-h)}{h^2}$$

converge for each value of x in (α, β) to the limit zero, for $h = 0$, except *that for an enumerable set of points G this is not known to be the case, then, provided that at each point x of G the expression*

$$\frac{f(x+h)-2f(x)+f(x-h)}{h}$$

converge to the limit zero, for $h=0$, the function $f(x)$ is a linear function in the whole interval (α, β).

It should be observed that the condition

$$\lim_{h=0} \frac{f(x+h)-2f(x)+f(x-h)}{h}=0$$

is certainly satisfied at any point x at which the differential coefficient $f'(x)$ exists and is finite.

To prove the theorem, let it be assumed that

$$f(x)-f(\alpha)-\frac{x-\alpha}{\beta-\alpha}\{f(\beta)-f(\alpha)\}$$

has a positive value p at some point x_1 interior to (α, β); and let

$$\phi(x, k)=f(x)-f(\alpha)-\frac{x-\alpha}{\beta-\alpha}\{f(\beta)-f(\alpha)\}+k(x-\alpha)^2,$$

where k is a positive number. We have

$$\phi(\alpha, k)=0,\ \phi(\beta, k)=k(\beta-\alpha)^2, \text{ and } \phi(x_1, k)=p+k(x_1-\alpha)^2;$$

and hence, provided

$$k<\frac{p}{(\beta-\alpha)^2-(x_1-\alpha)^2}=K,$$

the number $\phi(x_1, k)$ is greater than $\phi(\beta, k)$, and than $\phi(\alpha, k)$. We shall suppose k to be so chosen that this condition is satisfied; it then follows that $\phi(x, k)$ has a maximum between α and β. The absolute maximum value of $\phi(x, k)$ may be attained once, or a finite number of times, or an infinite number of times, in the interval (α, β).

The points x at which this maximum is attained have an upper extreme $\bar{x}(<\beta)$, which must itself be a point at which the maximum of $\phi(x, k)$ is attained, as is seen, in the case in which $\bar{x}$ is an upper limit, from the condition of continuity of the function. We have therefore

$$\phi(\bar{x}+h, k)-\phi(\bar{x}, k)\leqq 0, \text{ and } \phi(\bar{x}-h, k)-\phi(\bar{x}, k)\leqq 0,$$

if h be sufficiently small; from which we conclude that, in case

$$\lim_{h=0} \frac{\phi(\bar{x}+h, k)-2\phi(\bar{x}, k)+\phi(\bar{x}-h, k)}{h^2} \text{ exist,}$$

its value is $\leqq 0$. It follows that $\bar{x}$ must belong to G; because the value of this limit is $2k$, and therefore >0, for any point which does not belong to G. Since $\bar{x}$ is a point of G, we have

$$\lim_{h=0}\left\{\frac{\phi(\bar{x}+h, k)-\phi(\bar{x}, k)}{h}+\frac{\phi(\bar{x}-h, k)-\phi(\bar{x}, k)}{h}\right\}=0;$$

and since the two fractions have the same sign, it follows that

$$\lim_{h=0} \frac{\phi(\bar{x}+h, k) - \phi(\bar{x}, k)}{h} = 0, \text{ and } \lim_{h=0} \frac{\phi(\bar{x}-h, k) - \phi(\bar{x}, k)}{h} = 0.$$

From this result we deduce that

$$\lim_{h=0} \frac{f(\bar{x}+h) - f(\bar{x})}{h} = \lim_{h=0} \frac{f(\bar{x}-h) - f(\bar{x})}{-h} = \frac{f(\beta) - f(\alpha)}{\beta - \alpha} - 2k(\bar{x} - \alpha).$$

To each value of k in the interval $(0, K)$, there corresponds one value of $\bar{x}$, and it is impossible that the same value of $\bar{x}$ can correspond to two different values k_1, k_2 of k. For if this were the case, we should have

$$k_1(\bar{x} - \alpha) = k_2(\bar{x} - \alpha),$$

and therefore $k_1 = k_2$, since $\bar{x} > \alpha$. Now it is impossible that the set of points k interior to the interval $(0, K)$ can be such that to each such point there corresponds a distinct point $\bar{x}$ belonging to the enumerable set G. We conclude that it is impossible that

$$f(x) - f(\alpha) - \frac{x - \alpha}{\beta - \alpha}\{f(\beta) - f(\alpha)\}$$

can have a positive value p at any point x_1 of the interval (α, β); and it can be shewn in a similar manner that there can be no negative value of the same function in the interval. It follows that the function must everywhere be zero, and therefore that $f(x)$ is linear in the interval (α, β).

213. Let us suppose that a continuous function $f(x)$, defined for the interval (α, β), is such that, in every interior point of any sub-interval belonging to an everywhere-dense set of sub-intervals, the condition

$$\lim_{h=0} \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} = 0 \text{ is satisfied.}$$

It follows from the theorem of § 211, that in any one of the sub-intervals $f(x)$ is a linear function of x; and thus the value of $f(x)$ in a sub-interval (a_n, b_n) is a linear function $A_n x + B_n$. The set of sub-intervals is complementary to a closed set of points G which is non-dense in (α, β). In case this closed set G be enumerable, each interval (a_n, b_n) abuts at each end on another interval of the set; thus we may suppose that (a_n, b_n) abuts on $(a_{n'}, b_{n'})$ at the end b_n, so that $b_n = a_{n'}$.

Let us now assume that, at each point of G, the condition

$$\lim_{h=0} \frac{f(x+h) - 2f(x) + f(x-h)}{h} = 0 \text{ is satisfied.}$$

We have then, at the point $x = b_n = a_{n'}$, by applying this condition, and also the condition of continuity of $f(x)$,

$$A_n b_n + B_n = A_{n'} b_n + B_{n'}, \text{ and } A_{n'} - A_n = 0;$$

from which we deduce that $B_n = B_{n'}$, and thus that the linear functions

$$A_n x + B_n, \quad A_{n'} x + B_{n'}$$

are identical. Therefore it follows that, in case G be a non-dense enumerable set, the function $f(x)$ must be a linear function $Ax + B$ in the whole interval (α, β). This is, in fact, a particular case of the theorem of § 212.

The condition $\lim\limits_{h=0} \dfrac{f(x+h) - 2f(x) + f(x-h)}{h} = 0$ being certainly satisfied at any point x at which $f(x)$ has a finite differential coefficient, we therefore obtain the following theorem:—

If $f(x)$ be a continuous function possessing everywhere in the interval (α, β) a finite differential coefficient, and the function be linear in each one of an everywhere-dense set of intervals complementary to an enumerable closed set of points G, then $f(x)$ is a linear function in (α, β).

If the closed set of points G were unenumerable, the preceding reasoning would no longer be applicable, except that, at an isolated point of G, it would establish that the linear functions in the two intervals which abut on one another at the isolated point must be identical. Confining therefore our attention to the case in which G is a perfect set, we see that a continuous function possessing everywhere a finite differential coefficient may exist, which is linear in each sub-interval complementary to a non-dense perfect set of points contained in the interval for which the function is defined, and yet the function need not be linear in the whole interval.

The existence of such functions will be effectively established in Chapter V, where it will be shewn that they may be obtained by the integration of continuous functions which have an everywhere-dense set of lines of invariability.

OSCILLATING CONTINUOUS FUNCTIONS.

214. Let us suppose that the continuous function $f(x)$ has no lines of invariability in the interval (α, β), and that everywhere in this interval it has a finite differential coefficient. If within (α, β) there be a maximum or minimum of $f(x)$, then at such a point $f'(x)$, which exists and is finite, must be zero. If the maxima and minima in (α, β) be everywhere-dense, then $f'(x)$ vanishes at every point of the everywhere-dense set; and if $f'(x)$ were continuous throughout (α, β) it would follow that it was everywhere zero, which would be contrary to the hypothesis that (α, β) is not a line of invariability.

It follows from this that *if in an interval (α, β), which contains no lines of invariability of the continuous function $f(x)$, the differential coefficient $f'(x)$*

everywhere exists and is continuous, there must be in the interval an everywhere-dense set of sub-intervals in each of which the function is monotone.

We have further the following theorem:—

If $f(x)$ be continuous in (α, β), and have no lines of invariability, but have an everywhere-dense set of maxima and minima, there must be in the interval an everywhere-dense set of points at each of which $f'(x)$ either does not exist, or does exist and is discontinuous.

A continuous function $f(x)$, which in a given interval (α, β) has no lines of invariability, but has an everywhere-dense set of maxima and minima, is said to be a continuous function which is *everywhere-oscillating* in the interval (α, β). Such a function cannot have a differential coefficient which is continuous throughout the interval.

The continuous functions which are everywhere-oscillating in an interval may be divided into two classes.

(1) The function may be such that, if the constants l, m be properly chosen, the function $f(x) + lx + m$ is monotone in the interval. In this case $f(x)$ is expressible as the difference of two monotone functions, and thus belongs to the class of functions with limited total fluctuation. These functions may be said to be of the *first species*, or to be functions with *removable* oscillations.

(2) Such functions as do not belong to (1) may be said to be of the *second species*, or to be functions with *irremovable* oscillations.

In order to bring to light the essential distinction between the two classes of functions, as exhibited by the properties of their derivatives, we first of all remark that, if $D_+ f(x)$ have a positive lower limit c for all points x in the interval (α, β), then at each point $f(x+h) - f(x)$ is essentially positive for all positive values of h which are less than some number δ dependent on x; hence the function is monotone in the interval. The function would also be monotone in case the specified condition were that $D^+ f(x)$ has a negative upper limit for all values of x in (α, β).

Now suppose that $D_+ f(x)$ has a definite negative lower limit in (α, β); let this be $-c$, and consider the function $\phi(x) = f(x) + lx + m$, where $l > c$; we have then $D_+ \phi(x) = l + D_+ f(x) \geqq l - c$; hence the function $\phi(x)$ is monotone in (α, β). Thus $f(x)$ is expressible as the difference of the two monotone functions $\phi(x)$ and $lx + m$. Similarly, if we had taken the condition that $D^+ f(x)$ has a definite positive upper limit c, the function $f(x) + lx + m$, where $l < -c$, could be shewn to be monotone.

It is clear that instead of the linear function $lx + m$ we might have used any continuous differentiable function whose differential coefficient was $> c$, or $< -c$, throughout the interval, in the two cases.

The argument would have been unaltered if it had been assumed that there were a finite or infinite set of lines of invariability in (α, β).

It has thus been shewn that:—

If the continuous function $f(x)$ be such that either $D_+f(x)$ has a negative lower limit for all values of x in (α, β), or that $D^+f(x)$ has a positive upper limit, then all maxima and minima of the oscillating function $f(x)$ are removed by adding to $f(x)$ a properly chosen linear function, and thus the function is of the first species, and is of limited total fluctuation.

In particular, the conditions of the theorem are satisfied if the derivative, on one side, without necessarily having a definite value at any point, be such that for the whole interval it is numerically less than some fixed positive number.

A function, such that for a given interval,

$$|D^+f(x)|, \quad |D_+f(x)|, \quad |D^-f(x)|, \quad |D_-f(x)|$$

are all less than some fixed number, is said to be a function *with limited derivatives*. Such a function has a limited total fluctuation in the interval, and if it be everywhere-oscillating, it is of the first species.

A function with limited derivatives is necessarily a continuous function, but the converse does not hold.

In the general case, one of the derivatives on the right may at some or all of the points of the interval have indefinitely great values, this derivative being the same one for all such points.

If neither $D_+f(x)$ have a definite negative lower limit, nor $D^+f(x)$ have a definite positive upper limit in the interval, and the function be an everywhere-oscillating function, then it is of the second species.

215. Let us suppose that, for a set of points G, everywhere-dense in (a, b), the derivative of the continuous function $f(x)$ is infinite, but not of fixed sign, *i.e.* the derivatives at a point of G on the right and on the left exist, and are infinite, but of opposite signs. At any point x_0 of G, a neighbourhood can be found, containing x_0, such that for any point x in it $f(x)-f(x_0)$ is of fixed sign for the whole neighbourhood, and is never zero except when $x=x_0$; it follows that x_0 is a proper maximum or minimum of the function.

It will be shewn that, in any interval (α, β) contained in (a, b), there are an infinite number of points at which the function has the same value. Let ξ be a maximum point of $f(x)$ within (α, β), and let $(\xi-\eta, \xi+\epsilon)$ be the greatest interval enclosing ξ, for which $f(x)-f(\xi)$ is negative; suppose that the absolute minimum of the function for this interval is in $(\xi-\eta, \xi)$; taking a maximum point ξ_1 in the interval $(\xi, \xi+\epsilon)$, then in $(\xi-\eta, \xi)$ there is a point ξ_1' at which $f(\xi_1')=f(\xi_1)$, since $f(\xi_1)$ lies between the greatest and least values of the continuous function in $(\xi-\eta, \xi)$.

Now there is a maximum interval $(\xi_1 - \eta_1, \xi_1 + \epsilon_1)$ for the point ξ_1, and this lies within $(\xi, \xi + \epsilon)$; and in this interval we may as before find a maximum point ξ_2, such that a point ξ_2' also exists within the interval, for which $f(\xi_2) = f(\xi_2')$. There is also a point ξ_2'' in $(\xi - \eta, \xi)$, such that

$$f(\xi_2'') = f(\xi_2') = f(\xi_2).$$

We may proceed in this manner, until we find n points

$$\xi_{n-1},\ \xi'_{n-1}, \dots \xi_{n-1}^{(n-1)},$$

such that
$$f(\xi_{n-1}) = f(\xi'_{n-1}) = \dots = f(\xi_{n-1}^{(n-1)}).$$

Now let ξ_ω be a limiting point of $\xi_1, \xi_2, \xi_3, \dots \xi_n \dots$; and let ξ_ω' be a limiting point of $\xi_1', \xi_2', \dots$, and ξ_ω'' be a limiting point of $\xi_2'', \xi_3'', \dots$; then

$$f(\xi_\omega) = f(\xi_\omega') = f(\xi_\omega'') = \dots.$$

Thus the points $\xi_\omega, \xi_\omega', \xi_\omega'', \dots$ form an infinite set in (α, β) at which the functional values are the same.

The points $\xi_\omega, \xi_\omega', \xi_\omega'', \dots$ have a limiting point ξ_0 at which the functional value is the same as for the set itself; therefore

$$\frac{f(\xi_0) - f(\xi_\omega)}{\xi_0 - \xi_\omega} = \frac{f(\xi_0) - f(\xi_\omega')}{\xi_0 - \xi_\omega'} = \dots = 0;$$

hence at ξ_0 either the derivative is determinate and equal to zero, or else it is indeterminate with zero lying between its upper and lower limits. Thus it has been shewn that*:—

If the continuous function $f(x)$, have an everywhere-dense set of points at which the derivative is infinite but not of fixed sign, there is an everywhere-dense set of points at each of which the derivative is either indeterminate or else zero. Thus a continuous function cannot at all points have a derivative which is infinite and not of fixed sign.

If we apply the above theorem to the function $f(x) - cx$, where c is a prescribed constant, then, since $f(x) - cx$ has an infinite derivative at the same points as those for which $f(x)$ has an infinite derivative, we obtain the following theorem:—

If the continuous function $f(x)$ have at an everywhere-dense set of points a derivative which is infinite but not of fixed sign, there is an everywhere-dense set of points at each of which the derivative either has the prescribed value c, or is indeterminate, and such that c lies between its upper and lower limits.

* König, *Monatshefte f. Math. u. Physik*, vol. I. The above proof is that given by Schönflies, *Bericht*, p. 160.

GENERAL PROPERTIES OF DERIVATIVES.

216. A large number of properties of the derivatives of special classes of functions, chiefly belonging to the oscillating, or to the monotone, continuous functions, have been given by Dini and other writers; the most important of these will be given here.

The following general theorem, due* to W. H. Young, includes as a special case a theorem for continuous functions due to Brodén†.

The points at which one at least of the four derivatives of any given function is infinite, form an inner limiting set. The set of such points is accordingly of power c, when it contains a component dense-in-itself; and otherwise it is enumerable, or finite, or zero.

It follows that the set of such points is a set of the second category in case it be everywhere-dense.

Let x_0 be a point at which one of the four derivatives is infinite, it being immaterial whether the other derivatives are infinite or finite. A sequence $x_1, x_2, \dots x_n \dots$ converging to x_0, and on one side of it, can be found, which has the property that, corresponding to an arbitrarily large positive number σ, an integer m_1 can be found such that

$$\left|\frac{f(x_n)-f(x_0)}{x_n-x_0}\right| > \sigma, \quad \text{for } n > m_1;$$

further, m' can be chosen so great that

$$|x_n - x_0| < \frac{1}{\sigma}, \quad \text{for } n > m' \geqq m_1.$$

Let the intervals (x_0, x_n) be prolonged on the side beyond x_0, each being increased by $\frac{1}{\sigma-1}$ of its length; and the whole set of intervals so constructed for every point x of the set at which a derivative is infinite, may be called I_σ.

Let $\sigma_1, \sigma_2, \sigma_3 \dots$ be a set of values of σ which increase without limit; then the corresponding sets of intervals $I_{\sigma_1}, I_{\sigma_2}, \dots$ define an inner limiting set of points, to which all the points x of the given set belong; and it will be shewn that no other points belong to this inner limiting set. If possible let ξ be a point of the inner limiting set which does not belong to the given set of points at which a derivative is infinite. There is at least one interval of each of the sets $I_{\sigma_1}, I_{\sigma_2}, \dots$ such that ξ is an interior point of it; let such intervals be $\delta_1, \delta_2, \dots$, and let $\xi_1, \xi_2, \xi_3 \dots$ be points of the given set interior to these intervals. Let $\bar{\xi}_1, \bar{\xi}_2, \bar{\xi}_3 \dots$ be the end-points of the intervals on the sides of those intervals which were not lengthened. We have

$$\delta_\iota = \left(1 + \frac{1}{\sigma_\iota - 1}\right) |\xi_\iota - \bar{\xi}_\iota| < \frac{1}{\sigma_\iota - 1};$$

* *Arkiv för Matematik, Astronomi och Fysik*, vol. I, Stockholm, 1903.
† *Acta Univ. Lund.* vol. XXXIII, p. 81.

thus the points $\xi_1, \xi_2, \xi_3 \dots$ and also the points $\bar{\xi}_1, \bar{\xi}_2, \dots$ form a sequence of which ξ is the limit.

Since $\left|\dfrac{f(\xi_\iota) - f(\bar{\xi}_\iota)}{\xi_\iota - \bar{\xi}_\iota}\right| > \sigma_\iota$, therefore $|f(\xi_\iota) - f(\bar{\xi}_\iota)| > (\sigma_\iota - 1)\,\delta_\iota$.

Also a positive number A can be determined, such that for all values of ι,

$$\left|\frac{f(\xi) - f(\xi_\iota)}{\xi - \xi_\iota}\right| < A,$$

for otherwise ξ would be a point with an infinite derivative; and from this we see that

$$|f(\xi) - f(\xi_\iota)| < A\delta_\iota.$$

For a sufficiently great value of ι,

$$\sigma_\iota - 1 > A;$$

hence for such a value of ι,

$$|f(\xi) - f(\bar{\xi}_\iota)| > (\sigma_\iota - A - 1)\,\delta_\iota$$

and thus

$$\left|\frac{f(\xi) - f(\bar{\xi}_\iota)}{\xi - \bar{\xi}_\iota}\right| > \sigma_\iota - A - 1.$$

Now $\sigma_\iota - A - 1$ is arbitrarily large for a sufficiently great ι; hence, since ξ is the limiting point of the sequence $\{\bar{\xi}_\iota\}$, there is an infinite derivative at ξ, which is contrary to the hypothesis made; therefore the points of the given set constitute the inner limiting set which has been defined.

217. If x_1 be a point of the interval (a, b) in which $f(x)$ is defined, the function $\dfrac{f(x) - f(x_1)}{x - x_1}$ for points x such that $x_1 < x \leqq b$ may be called the incrementary ratio at x_1 on the right; and in case $f(x)$ be a continuous function, this incrementary ratio is also continuous at every point of its domain. This incrementary function has an upper and a lower limit for its whole domain $(x_1 < x \leqq b)$; and these upper and lower limits may be denoted by $U(x_1)$, $L(x_1)$, and either of them may be finite or infinite; however $U(x_1)$ can only be infinite with the positive sign, and $L(x_1)$ only with the negative sign. $U(x_1)$, $L(x_1)$ being regarded as functions of x_1, defined for every point of (a, b) except the point b, the function $U(x_1)$ has a finite or infinite upper limit for its whole domain, which we denote by U; and the function $L(x_1)$ has a finite or infinite lower limit for its whole domain, which we may denote by L. There exist therefore two finite numbers U, L, which may have the improper values $+\infty$, $-\infty$ respectively, such that

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1},$$

for every pair of values of x_1, x_2, where $x_2 > x_1$, always lies between them, or is equal to one of them.

The incrementary ratio on the left of a point can be defined in a similar manner; and we thus define two functions $U'(x_1)$, $L'(x_1)$ at x_1, as the upper and lower limits of these incrementary ratios.

It is easily seen that U', the upper limit of $U'(x)$ in the interval (a, b), is identical with U, and that L' the lower limit of $L'(x)$ is identical with L. Thus U, L are the upper and lower limits of

$$\frac{f(x_2)-f(x_1)}{x_2-x_1},$$

for every possible pair of points (x_1, x_2) in the interval (a, b).

218. Let $f(x)$ be continuous in the interval (a, b), and let U and L be the upper and lower limits of the incrementary ratios above defined. Take (α, β) any interval in (a, b), and consider the function

$$\phi(x) = f(x) - f(\alpha) - \frac{x-\alpha}{\beta-\alpha}[f(\beta)-f(\alpha)].$$

Since $\phi(\alpha) = 0$, $\phi(\beta) = 0$, unless $\phi(x)$ be constant through (α, β) there must be within (α, β) a maximum or minimum of (α, β); and thus at least one point x_1 exists within (α, β) such that

$$\phi(x_1 \pm h) - \phi(x_1) \leqq 0,$$

for all sufficiently small values of h, or else

$$\phi(x_1 \pm h) - \phi(x_1) \geqq 0,$$

for all sufficiently small values of h. At such a point

$$\frac{f(x_1+h)-f(x_1)}{h} \leqq \frac{f(\beta)-f(\alpha)}{\beta-\alpha},$$

and $$\frac{f(x_1-h)-f(x_1)}{-h} \geqq \frac{f(\beta)-f(\alpha)}{\beta-\alpha},$$

or else $$\frac{f(x_1+h)-f(x_1)}{h} \geqq \frac{f(\beta)-f(\alpha)}{\beta-\alpha},$$

and $$\frac{f(x_1-h)-f(x_1)}{-h} \leqq \frac{f(\beta)-f(\alpha)}{\beta-\alpha}.$$

If $\phi(x)$ have an infinite number of maxima and minima in (α, β), there are in (α, β) an infinite number of points at which the first of these conditions for $\phi(x)$ holds, and also an infinite number at which the second holds. If there be only a finite number of maxima and minima of $\phi(x)$ in (α, β), then this interval can be divided into a number of portions in each of which the function $\phi(x)$ is monotone; and in any one of these portions either

$$\frac{f(x \pm h)-f(x)}{\pm h} \leqq \frac{f(\beta)-f(\alpha)}{\beta-\alpha},$$

at all points within the sub-interval, or else

$$\geqq \frac{f(\beta)-f(\alpha)}{\beta-\alpha},$$

for every x within the portion, and for sufficiently small values of h. Now let U, L be the upper and lower limits of $\frac{f(x_2)-f(x_1)}{x_2-x_1}$ in (a, b), then

$$\frac{f(\beta)-f(\alpha)}{\beta-\alpha} \text{ and } \frac{f(x\pm h)-f(x)}{\pm h}$$

lie between U and L. Thus, in every interval (α, β) contained in (a, b), in which $\phi(x)$ has an infinite number of maxima and minima, there are (1) an infinity of points x for which $\frac{f(x+h)-f(x)}{h}$, for all sufficiently small values of h, lies between L and $\frac{f(\beta)-f(\alpha)}{\beta-\alpha}$; and (2) an infinity of points for which the same is true of $\frac{f(x-h)-f(x)}{-h}$; (3) an infinity of points for which $\frac{f(x+h)-f(x)}{h}$, for all sufficiently small values of h, lies between U and $\frac{f(\beta)-f(\alpha)}{\beta-\alpha}$; and (4) an infinity of points for which the same is true of $\frac{f(x-h)-f(x)}{-h}$.

In case $f(x)-f(\alpha)-\frac{x-\alpha}{\beta-\alpha}[f(\beta)-f(\alpha)]$ have only a finite number of maxima and minima in (α, β), there are in (α, β) finite intervals such that all the points in one of them belong to both the sets (1) and (2), and also finite intervals in which all the points belong to both the sets (3) and (4); each of these sets of intervals is finite, and an interval of one set is followed by one of the other set.

The number L being the lower limit of the function $L(x)$ in the interval (a, b), there exists a point x_1 such that L is the lower limit of the values of $L(x)$ in any arbitrarily small neighbourhood of x_1; and it follows that in such neighbourhood of x_1 there are points ξ such that $\frac{f(\xi+h)-f(\xi)}{h}$, for an infinity of values of h, differs from L by less than a prescribed positive number ϵ. Therefore there are in (a, b) an infinity of pairs of points (α, β) one of which is arbitrarily near x_1, such that $\frac{f(\beta)-f(\alpha)}{\beta-\alpha}$ differs from L by less than ϵ.

Similarly it may be shewn that in (a, b), there are an infinity of pairs of points (α, β) such that $\frac{f(\beta)-f(\alpha)}{\beta-\alpha}$ differs from U by less than the prescribed number ϵ.

If U or L be infinite, there exists an infinity of pairs of points such that $\frac{f(\beta)-f(\alpha)}{\beta-\alpha}$ is arithmetically greater than a prescribed number c, and has the same sign as the infinite U or L. We can consequently choose the interval (α, β) so that

$$\frac{f(\beta)-f(\alpha)}{\beta-\alpha}=L+\eta,$$

or else so that

$$\frac{f(\beta)-f(\alpha)}{\beta-\alpha}=U-\eta,$$

where $\eta<\epsilon$, provided U and L are finite. If one or both of U, L be infinite, (α, β) can be so chosen that $\frac{f(\beta)-f(\alpha)}{\beta-\alpha}$ has the same sign as U or as L, and is arithmetically greater than a prescribed positive number c.

We have now obtained the following results:—

If $f(x)$ be a continuous function, and (a, b) be the whole or a part of its domain, to which U and L correspond, then (1) *if L be finite, there exists in (a, b) an infinity of points for which both $D^+(x)$, $D_+(x)$ each lie between L and $L+\epsilon$, where ϵ is an arbitrarily prescribed positive number; and at these points $D^+(x)$, $D_+(x)$ are either equal, in which case a derivative on the right exists, or else they differ from one another by less than ϵ:* (2) *if U be finite there exists in (a, b) an infinity of points for which $D^+(x)$, $D_+(x)$ each lie between U and $U-\epsilon$; and at these points there exist derivatives on the right, or else $D^+(x)$, $D_+(x)$ differ from one another by less than ϵ;* (3) *if U or L be infinite there exists an infinity of points at which $D^+(x)$, $D_+(x)$ are both numerically greater than an arbitrarily great number c, and have the same sign as the U or L which is infinite. A similar statement holds as regards the derivatives on the left.*

The above is true irrespectively of the number of the maxima and minima of $f(x)$; but if $f(x)$ have in (a, b) only a finite number of maxima and minima and if the same be true of all the functions $f(x)-lx-m$, obtained by the addition of a linear function, then there exist in (a, b) finite sub-intervals such that at all points in one of them the above statements hold both as regards the derivatives on the right and as regards those on the left. The numbers U and L correspond in each case to the particular sub-interval.

It will be observed that the theorem does not assert the necessity of the existence of points at which a determinate derivative on the right or on the left exists, but it states that there are in every sub-interval points at which the difference between the upper and lower derivatives on one side is less than a prescribed arbitrarily small number, or else at which both such derivatives are arithmetically greater than an arbitrarily fixed large number. There are therefore certainly points in every sub-interval at which there is, so to speak, an arbitrarily near degree of approximation to the existence of a finite or infinite derivative on the right, and also points at which the same is true as regards derivatives on the left.

219. It will now be shewn* that, *for a continuous function, of which* (a, b) *is the whole or a part of its domain, the upper limit of each of the four derivatives* $D^+f(x)$, $D_+f(x)$, $D^-f(x)$, $D_-f(x)$ *for all values of* x *in* (a, b) *is* U *the upper limit of the incrementary function in* (a, b), *and that the lower limit of the four functions is* L. *If* U *and* L *be both finite the function belongs to the class of functions with limited derivatives.*

A function with limited derivatives accordingly satisfies the condition, that for every pair of points x_1, x_2, $|f(x_1) - f(x_2)| < k|x_1 - x_2|$, where k is a fixed positive number. It has been pointed out in § 214, that such a function belongs to the class of functions of limited total fluctuation.

It is clear that the upper limit of each of the functions $D^+f(x)$, $D_+f(x)$, $D^-f(x)$, $D_-f(x)$ is a number which cannot be greater than U. Now since it has been shewn that points exist in (a, b) such that, if ϵ be an arbitrarily prescribed number, both $D^+f(x)$, $D_+f(x)$ differ from U by less than ϵ, when U is finite, and are arbitrarily great if U is $+\infty$, it follows that U is in either case the upper limit of $D^+f(x)$, $D_+f(x)$. In a similar manner it can be shewn that U is the upper limit of both $D^-f(x)$, $D_-f(x)$. The proof that L is the common lower limit of the four functions is exactly similar.

Each of the four expressions $D^+f(x)$, $D_+f(x)$, $D^-f(x)$, $D_-f(x)$ may be regarded as a function defined for the whole domain of $f(x)$ except at one of the end-points; but in this case we have to extend the ordinary definition of a function so far as to admit infinite functional values, instead of only infinite functional limits as in the case of an ordinary function. It will be convenient to say that, at a point at which one of the above functions is infinite, it is also continuous, provided the functional limits of the function in question, on either side, are definitely infinite and of the same sign as the functional value at the point.

If, at any point x_0, interior to (a, b), one of the above functions, say $D^+f(x)$, be continuous, then at that point the other three functions are also continuous, and are equal in value to $D^+f(x_0)$, and thus there exists at x_0, a differential coefficient.

To prove this, take any interval $(x_0 - \epsilon, x_0 + \epsilon)$; then all four functions have in this interval the same upper limits, and also the same lower limits. If $D^+f(x_0)$ be finite, the upper and lower limits of $D^+f(x)$ in $(x_0 - \epsilon, x_0 + \epsilon)$ each differ from $D^+f(x_0)$ by less than a number η which depends on ϵ in such a way that, as ϵ is indefinitely diminished to the limit zero, η also diminishes to the limit zero. Since all four functions have the same upper

* Du Bois Reymond, *Math. Ann.* vol. XVI, p. 119, also Scheeffer, *Acta Mathematica*, vol. V, p. 190.

limit and the same lower limit in $(x_0 - \epsilon, x_0 + \epsilon)$, the upper and the lower limits of each differ from $D^+f(x_0)$ by less than η, and η can be made as small as we please by taking ϵ small enough. It follows that all four functions are continuous at x_0, and that all four at x_0 are equal to $D^+f(x_0)$; and thus there exists a differential coefficient at x_0.

In case $D^+f(x_0)$ is $+\infty$, ϵ can be so chosen that in $(x_0 - \epsilon, x_0 + \epsilon)$, $D^+f(x)$ is everywhere greater than an arbitrarily large chosen number c, and the upper and lower limits of each of the four functions are then greater than c; by taking a succession of values of c which increases indefinitely, and considering the corresponding sequence of values of ϵ which converges to zero, we see that each of the functions $D_+f(x)$, $D^-f(x)$, $D_-f(x)$ is infinite at x_0, and is continuous in the extended sense of the term, at that point; there is then a differential coefficient at x_0 which is infinite and of definite sign.

It follows that, *if it be known that any one of the four derivatives is everywhere continuous in an interval, there exists everywhere in the interval a differential coefficient in the ordinary sense of the term.*

220. *The derivatives $D^+f(x)$, $D_+f(x)$ of a continuous function are at any point x_0, such that $a \leqq x_0 < b$, either both continuous on the right, or both of them have a discontinuity of the second kind on the right; but they cannot have ordinary discontinuities on the right.*

A similar statement holds as regards the continuity or discontinuity of $D^-f(x)$, $D_-f(x)$ on the left.

Suppose that $D^+f(x)$ has at the point x_0, a limit λ at x_0 on the right; then if δ be a prescribed positive number, an interval $(x_0, x_0 + \epsilon)$ can be found, such that $D^+f(x)$, for every point of this interval, except x_0, lies between $\lambda + \delta$, and $\lambda - \delta$. The upper and lower limits of each of the four derivatives $D^+f(x)$, $D_+f(x)$, $D^-f(x)$, $D_-f(x)$ for any interval $(x_0 + \epsilon_1, x_0 + \epsilon)$, where $\epsilon_1 < \epsilon$, must all lie between the values $\lambda + \delta$, $\lambda - \delta$; hence the upper and lower limits of $D^-f(x)$ for the interval $(x_0, x_0 + \epsilon)$, lie between these same values, the function $D^-f(x)$ being regarded as undefined at the point x_0; and these upper and lower limits of $D^-f(x)$ are the same as those of $D^+f(x)$, $D_+f(x)$ for $(x_0, x_0 + \epsilon)$, the point x_0 being included. It follows that $D^+f(x_0)$, $D_+f(x_0)$ both lie between $\lambda + \delta$ and $\lambda - \delta$; and as this holds for every value of δ, we must have

$$D^+f(x_0) = D_+f(x_0) = \lambda = \lambda';$$

where λ' denotes the limit of $D_+f(x)$ at x_0 on the right; and thus $D^+f(x)$, $D_+f(x)$ are both continuous at x_0 on the right. If $\lambda = +\infty$, then in the interval $(x_0, x_0 + \epsilon)$ at every point except x_0, $D^+f(x) > c$, where c is an arbitrarily chosen number on which ϵ depends; the argument then proceeds as before.

221. *If a continuous function $f(x)$ be such that, of the functions $\phi(x) = f(x) - lx$, where l has all values, none, or only a finite number, or only an infinite set which does not fill any interval, however small, have more than a finite number of maxima and minima in sufficiently small neighbourhoods on the right of x_0, then the function $f(x)$ has a derivative on the right at x_0, either finite or infinite.*

If $f(x) - lx$ have only a finite number of maxima and minima in $(x_0, x_0 + \alpha)$, an interval $(x_0, x_0 + \epsilon_1)$ can be found, in which the function is monotone. Since then, at every point $x_0 + h$ in such an interval,

$$f(x_0 + h) - f(x_0) - lh \geqq 0,$$

or else at every point

$$f(x_0 + h) - f(x_0) - lh \leqq 0,$$

it follows that $D^+f(x_0)$, $D_+f(x_0)$ are either both $\geqq l$, or else both $\leqq l$, and this holds for a set of values of l, everywhere-dense in any interval whatever. If $D^+f(x_0)$, $D_+f(x_0)$ were unequal, we could find a value of l which is between the two and is not one of the exceptional values of l, and this would be contrary to what has been proved; hence $D^+f(x_0) = D_+f(x_0)$, or $f(x)$ possesses a derivative $d(x_0)$ on the right at x_0. This derivative is finite if some of the functions $f(x) - lx$ increase and others decrease on the right of x_0; otherwise it is infinite.

For any point x within $(x_0, x_0 + \epsilon_1)$, we have, for a sufficiently small value of h,

$$\frac{f(x \pm h) - f(x)}{\pm h} \geqq l,$$

if l be less than the derivative at x_0 on the right, and

$$\frac{f(x \pm h) - f(x)}{\pm h} \leqq l,$$

if l be greater than the same derivative $d(x_0)$. If $d(x_0)$ be finite, it follows that $\dfrac{f(x \pm h) - f(x)}{\pm h}$ lies between $d(x_0) + \sigma$, $d(x_0) - \sigma$, where σ is an arbitrarily small number, provided $x - x_0$ and h be sufficiently small. If $d(x_0)$ is ∞ or $-\infty$, $\dfrac{f(x \pm h) - f(x)}{\pm h}$ lies between c and $+\infty$, or $-c$ and $-\infty$, where c is an arbitrarily great positive number. In either case therefore the four functions $D^+f(x)$, $D_+f(x)$, $D^-f(x)$, $D_-f(x)$ for points within $(x_0, x_0 + \epsilon_1)$ have the same limit at x_0, viz. $d(x_0)$. It follows, as a particular case, that if, under the conditions of the above theorem, a derivative on the right exist at all points in an arbitrarily small neighbourhood on the right of x_0, the derivative on the right at x_0 is a continuous function on the right. If derivatives on the left exist at all points in a neighbourhood of x_0 on the right, with the possible exception of x_0 itself, these derivatives have, at the point x_0, the derivative $d(x_0)$ as their limit.

222. *A continuous function $f(x)$ cannot have, at every point of a whole interval, a single-valued derivative on the right, which is everywhere infinite and of the same sign.*

For if $f(x)$ had this property in an interval (α, β), so also would

$$f(x)-f(\alpha)-\frac{x-\alpha}{\beta-\alpha}[f(\beta)-f(\alpha)],$$

and this function necessarily has a maximum or minimum within (α, β), which is contrary to the condition that it has a derivative on the right which is always of the same sign; for this involves the condition that the function must constantly increase as x increases from α to β.

Let us now suppose that the continuous function $f(x)$ has at all points of (a, b) single-valued derivatives on the right (finite or infinite), such that, in a part (α, β) of (a, b), this derivative is continuous at least on one side; *the function $f(x)$ is then such that, at an infinite number of points, it possesses an ordinary differential coefficient.*

The derivative $d(x)$ on the right cannot at all points of (α, β) be infinite. For if we take a point x_0 such that it is continuous on one side, in the extended sense of the term explained in § 219, then if it were everywhere infinite, its sign at all points in an interval on the one side of x_0 would be the same; but it has been shewn to be impossible that, everywhere in any interval, $d(x)$ should be infinite and of constant sign. It follows that there are points in the neighbourhood of x_0 at which $d(x)$ is finite. If x_1 be such a point in (α, β), then, since $d(x)$ is continuous on one side at x_1, an interval can be found at all points of which $d(x)$ is finite, and also continuous on one side. If (α_1, β_1) be such an interval in (α, β), then since $d(x)$ is everywhere finite in it, and continuous on one side at least, it is a point-wise discontinuous function, if it be not continuous in (α_1, β_1); and there must therefore be an infinity of points in (α_1, β_1) at which $d(x)$ is continuous. At such points, in accordance with § 219, $f(x)$ has a differential coefficient.

223. If we now collect the results obtained in § 221 and § 222, we can state the following general theorem, applicable to functions which are in general monotone, and also to a certain class of everywhere-oscillating functions.

If a continuous function $f(x)$ in a whole interval (a, b) be such that, corresponding to each point x_0, there be of the functions $f(x)-lx$ at most only a finite number, or an infinite set for which the values of l do not fill any interval, which, in an arbitrarily small neighbourhood on the right of x_0, contains an infinite number of maxima and minima, and if the same condition be true as regards an arbitrary small neighbourhood on the left of x_0, then the function has at every point of (a, b) a definite derivative on the right, and also a definite derivative on the left, and there is an everywhere-dense set of points at which there is a differential coefficient.

In every part of (a, b) there are finite intervals, in each one of which the derivatives on the right and those on the left are both definite and finite, and such that each of them is, in the interval to which it belongs, a point-wise discontinuous, or else a continuous, function.

As regards everywhere-oscillating functions, the following remarks may be made.

If a continuous function have in every neighbourhood on the right of a point x_0, an infinite number of maxima and minima, there are in such neighbourhoods an infinity of points at which the derivatives on the right are negative or zero, and an infinity of points in which these derivatives are positive or zero. It follows from this, that none of the derivatives at a point x on the right of x_0 can have a definite limit as x approaches the limit x_0, unless such limit be zero.

In particular, if at all such points x, definite derivatives on the right and on the left exist, these derivatives cannot be continuous at x_0, unless the derivatives at x_0 are both zero.

If at the point x_0, and at every point in a neighbourhood of x_0 which contains an infinite number of maxima and minima, a differential coefficient exist, which is continuous at x_0, this differential coefficient must be zero at x_0 and at an infinity of points in the neighbourhood of x_0, and must therefore itself have an infinite number of maxima and minima in the neighbourhood of x_0.

If a function $f(x)$, which has an infinite number of maxima and minima in the neighbourhood of x_0, have at x_0, and in its neighbourhood, differential coefficients of any number of orders, then they are all functions with an infinite number of maxima and minima in the neighbourhood of x_0, and all of them vanish at x_0, except that the one of highest order may be discontinuous at x_0, not then necessarily vanishing at that point. If differential coefficients of all orders exist, they must all vanish at x_0; and such a function is incapable of expansion in powers of $x - x_0$ in the neighbourhood of x_0. An example* of a function of this kind is

$$x^2 + e^{-\frac{1}{(x-x_0)^2}} \sin \frac{1}{x - x_0}.$$

FUNCTIONS WITH ONE DERIVATIVE ASSIGNED.

224. *If two functions, defined for a given interval, have each limited derivatives, and if the two functions have one of their four derivatives, say the upper derivative on the right, equal to one another at every point which does not belong to a set of points E of measure zero, then the two functions differ from one another by a constant, the same for the whole interval.*

* Dini, *Grundlagen*, p. 314.

This theorem* differs from that of § 206, in the respect that the functions are restricted to be such continuous functions as have limited derivatives; it is however more general, in that E is not restricted to be enumerable.

Let the points of E be enclosed in the interiors of intervals of a set, of which the total length has the arbitrarily small value ϵ. To each point P of E there corresponds an interval PP', where PP' is that part of the interval of the set that encloses P which is on the right of P; these intervals PP' may be denoted by δ'. If the two functions $f_1(x)$, $f_2(x)$ are such that, at a point x_1, $D^+f_1(x_1) = D^+f_2(x_1)$, it has been seen in § 206, that $D^+f(x_1) \geqq 0$, $D_+f(x_1) \leqq 0$, where $f(x)$ denotes $f_1(x) - f_2(x)$. Since $f(x)$ is continuous at x_1, it follows that there is a set of points $x_1 + h$ on the right of x_1, such that $|f(x_1 + h) - f(x_1)| \leqq \epsilon h$; if we suppose h to have the greatest value for which this holds, the interval $(x_1, x_1 + h)$ is an interval on the right of x_1, and such intervals may be denoted by δ.

Let ξ be any point such that $a < \xi \leqq b$, and consider the interval (a, ξ). From the point a lay off an interval δ or δ', according as a is not, or is, a point of E; from the end of this interval lay off another interval δ or δ', as the case may be. Proceeding in this manner, we may either reach the point ξ, after taking a finite number of intervals, or else we obtain an infinite set of intervals, the end-points of which have a limiting point P_ω, which may, or may not, coincide with ξ. In the latter case, we commence again to lay off intervals on the right of P_ω, until we either reach ξ, or else until another limiting point $P_{\omega 2}$ is obtained as the limit of a sequence of end-points. Proceeding in this manner, the point ξ is certainly reached as the end of an ordered sequence of intervals corresponding to a set of ordinal numbers which comes to an end before some number of the first or of the second class. The set of points not interior to the intervals is a closed enumerable set. We can now find $f(\xi) - f(a)$ as the sum, or limiting sum, of the differences of the functional values at the end-points of the intervals which have been defined, and each of which is either a δ, or a δ'. It is clear that

$$|f(\xi) - f(a)| \leqq \epsilon \Sigma \delta + A \Sigma \delta' < \epsilon (\xi - a + A),$$

where the summations refer to those of the intervals δ, δ' which have been employed in the construction; and A denotes the finite upper limit of $\left|\frac{f(x_1) - f(x_2)}{x_1 - x_2}\right|$ for every pair of points x_1, x_2 in the interval (a, ξ), and which is identical with the upper limit of the absolute value of the derivatives of $f(x)$ in the interval. Since ϵ is arbitrarily small, it follows that $f(\xi) = f(a)$, and therefore $f_1(\xi) - f_2(\xi) = f_1(a) - f_2(a)$; thus the theorem has been established.

* Lebesgue's *Leçons sur l'intégration*, p. 79.

THE CONSTRUCTION OF CONTINUOUS FUNCTIONS.

225. One of the most fruitful methods of obtaining continuous functions which exhibit various peculiarities as regards the existence or non-existence of differential coefficients at all the points, or at sets of points of their domain, consists of defining the functions by means of series specially constructed with a view to the purpose on hand; this method will be explained and illustrated in Chapter VI. Brodén, Köpcke and others have however given direct constructions for continuous functions, which illustrate various possibilities in relation to the existence and properties of derivatives.

The method employed* by Brodén is that of defining a continuous function in the domain (a, b), as the function obtained by *extension* (see § 191) of a function defined for an enumerable everywhere-dense set of points in (a, b), the primary points. A continuous function is entirely determinate when the functional values at such a primary set of points have been assigned. The necessary and sufficient condition that a function defined for the primary set should, by extension to the domain (a, b), give a function which is continuous in that domain, is that the primary function should be *uniformly continuous* with respect to the unclosed primary domain. To prove this, let $\{\xi\}$ denote the set of primary points, and $\{x\}$ the set of secondary points; then the condition that the function $f(\xi)$ may be uniformly continuous with respect to the domain $\{\xi\}$, is that, if ξ_1 be any point of $\{\xi\}$, and if η be a prescribed arbitrarily small number, the condition $|f(\xi)-f(\xi_1)| < \eta$ be satisfied at all points ξ which are such that $|\xi - \xi_1| < \epsilon$, where ϵ is a number dependent on η, but the same for all points ξ_1 of $\{\xi\}$. Now assuming that this condition is satisfied let x_1 be a secondary point, and let $\xi_1, \xi_2, \dots \xi_n \dots, \xi_1', \xi_2', \dots \xi_n' \dots$ be any two sequences of primary points each of which has x_1 as its limit; we have to shew that each of the sequences

$$f(\xi_1), f(\xi_2), \dots f(\xi_n) \dots,$$
$$f(\xi_1'), f(\xi_2'), \dots f(\xi_n') \dots$$

converges to the same number, which will then be the single functional value $f(x_1)$. Enclose x_1 in the interval $(x_1 - \frac{1}{2}\epsilon, x_1 + \frac{1}{2}\epsilon)$; then, from and after some particular value of n, all the points of both sequences of values of ξ lie within this neighbourhood. Let this value of n be m, then

$$|f(\xi_m) - f(\xi_{m+r})| < \eta,$$

for all positive integral values of r; hence the first sequence of functional values is convergent, since η is arbitrary; and similarly the second is also convergent. Also for every η there is a definite m such that

$$|f(\xi_{m+r}) - f(\xi'_{m+r})| < \eta;$$

* *Crelle's Journal*, vol. CXVIII; see also *Acta Univ. Lund.* vol. XXXIII.

hence the two convergent sequences have the same limit, and this limit defines $f(x_1)$. We have now to shew that the single-valued function so defined is continuous. We have

$$|f(x_1)-f(\xi_1)|<\eta, \text{ provided } |x_1-\xi_1|<\tfrac{1}{2}\epsilon,$$

and

$$|f(x_2)-f(\xi_2)|<\eta, \text{ provided } |x_2-\xi_2|<\tfrac{1}{2}\epsilon;$$

also

$$|f(\xi_2)-f(\xi_1)|<\eta, \text{ provided } |\xi_2-\xi_1|<\epsilon.$$

Hence it follows that $|f(x_2)-f(x_1)|<3\eta$,

and this holds provided $|x_2-x_1|<2\epsilon$,

for ξ_1, ξ_2 can be taken to be between x_1 and x_2; and therefore $f(x)$ is continuous at x_1, since 3η is at our choice. The extended $f(x)$ is also easily seen to be continuous at any primary point. It has now been proved that the condition of uniform continuity is sufficient; that it is necessary follows from the theorem of § 175.

The derivatives at any point depend only on the functional values at the primary points in the neighbourhood of the point. For let x_1 be any point, and consider the limit of $\dfrac{f(x)-f(x_1)}{x-x_1}$, when x has any sequence of values which converge to x_1. A set of primary values of x can always be found, such that the ratio converges to the same limit, when x has the values of this sequence of primary points, as for the prescribed sequence consisting of secondary points, or of both primary and secondary points. For a primary point ξ can be found, corresponding to x, such that

$$\left|\frac{f(x)-f(x_1)}{x-x_1}-\frac{f(\xi)-f(x_1)}{\xi-x_1}\right|<\delta,$$

where δ is an arbitrarily small number. This follows from the fact that

$$\frac{f(x)-f(x_1)}{x-x_1}$$

is a continuous function of x at every point except x_1.

226. In order to construct monotone continuous functions, the values of the function are first assigned at the end-points a, b of the interval, then at two points x_0, x_1, where $x_0<x_1$; then at four points x_{00}, x_{01}, x_{10}, x_{11}, where

$$x_{00}<x_0,\quad x_0<x_{01}<x_{10}<x_1, \text{ and } x_1<x_{11};$$

afterwards at eight points

$$x_{000},\ x_{001},\ x_{010},\ x_{011},\ x_{100},\ x_{101},\ x_{110},\ x_{111},\ \&c.$$

lying in the successive intervals measured from left to right, into which (a, b) was divided by the four points; and so on. The function may then be regarded as the limit of a sequence of continuous functions, each of which is representable as a polygon obtained by joining the end-points of ordinates which represent the functional values that have been assigned at any stage of the process.

In this manner Brodén has constructed a monotone continuous function $f(x)$, which is such that it has derivatives on the right, and on the left, which are everywhere definite, finite and different from zero; and such that a definite differential coefficient everywhere exists, except at the everywhere-dense enumerable set of primary points.

He has also constructed a monotone function $f(x)$ which is such that at the everywhere-dense enumerable set of primary points, the derivative on the left exists, and is zero, and the derivative on the right exists and is positive; for an unenumerable everywhere-dense set of points there is a differential coefficient everywhere zero, and for another such set of points, there is no definite derivative on the left, but there is a positive one on the right.

A third case is the following:—

$f(x)$ is continuous, monotone and increasing; at an everywhere-dense enumerable set of points the derivative on the left is zero, and that on the right is $+\infty$; for an everywhere-dense unenumerable set, both derivatives exist and are positive; for another such set both exist and are zero; for a third such set, both derivatives exist and are $+\infty$; for a fourth such set, neither derivative exists; for a fifth such set, the derivative on the left is zero, and that on the right is indefinite, but has zero for its lower limit; for a sixth such set, the derivative on the right is $+\infty$, but that on the left is indefinite, with $+\infty$ for its upper limit.

227. For the construction of everywhere-oscillating continuous functions it is more convenient to successively assign the functional values at sets of points proceeding by powers of 3 instead of 2 as in the case of monotone functions. In this manner Brodén has constructed such a function $f(x)$, which has the following properties:—

At an everywhere-dense enumerable set of points, the derivative on the left exists, and is positive; that on the right exists, and is negative (or the reverse), this set corresponding to maxima and minima of the function; for a certain unenumerable everywhere-dense set, there is a differential coefficient everywhere of the same sign; and for another such set, there is a differential coefficient which is zero; for a third such set, one or both of the derivatives are indefinite.

Köpcke* has given the first example of a function which is everywhere-oscillating and yet has at every point a definite differential coefficient, thus confirming the conjecture of Dini that such functions can exist; and Brodén† has also constructed such a function. A general theory of such functions has

* *Math. Ann.* vol. XXIX, p. 123; vol. XXXIV, p. 161; vol. XXXV, p. 104. See also Pereno, *Giorn. di Mat.* vol. XXXV, p. 132.

† *Stockholm Vet. Ak. Öfv.*, 1900, pp. 423 and 743.

been given* by Schönflies. The method adopted by Köpcke is to construct the function as the limit of a succession of polygons of which the sides are circular arcs. Everywhere-oscillating functions have also been studied by Steinitz†. A detailed account of all the special cases treated of by these writers would require a large amount of space; reference can therefore only be made to the original memoirs. A simplification of Köpcke's construction, due to Pereno, will be given in Chapter VI.

228. A function $f(x)$ which is of such a character that it can be represented approximately by a graph, which exhibits all the peculiarities of the function, so that $y=f(x)$ is the equation of a "curve," in the ordinary sense of the term, must satisfy the following three conditions:—

(1) The function must be continuous everywhere, with the possible exception of a finite number of points, at which it may have ordinary discontinuities.

(2) It must be differentiable, except that there may be a finite number of points at which no differential coefficients exist, but at which definite derivatives on the right and on the left exist.

(3) It can have only a finite number of maxima and minima; and the same must hold of every function obtainable by the addition of a linear function to the one in question. This condition may be expressed in the form, that the function must be in general monotone with reference to every possible axis which may be employed for the measurement of abscissae.

A function which satisfies these conditions may be characterised‡ as an *ordinary* function. As has been already indicated, there exist functions which satisfy the conditions (1) and (3), but do not satisfy the condition (2). Again, there exist functions which satisfy the conditions (1) and (2), but not the condition (3).

FUNCTIONS OF TWO OR MORE VARIABLES.

229. An association of n numbers $(a_1, a_2, \dots a_n)$ being considered to represent a point in n-dimensional space, any set of such points, whether continuous or not, may be taken as the domain of a set $(x_1, x_2, \dots x_n)$ of n independent variables. When $|x_1|, |x_2|, \dots |x_n|$ are, for all points of the domain, all less than some fixed positive number, the domain is said to be *limited*.

A function $f(x_1, x_2, \dots x_n)$, as in the case of a domain of one dimension, is defined by a set of rules from which a single number, the functional value,

* *Math. Ann.* vol. LIV; also his *Bericht*, p. 164.

† *Math. Annalen*, vol. LII.

‡ Du Bois Reymond, *Crelle's Journal*, vol. LXXIX, p. 32.

can be arithmetically determined for any prescribed point of the given domain. If, for every point of the domain, $|f(x_1, x_2, \dots x_n)|$ be less than some fixed positive number, the function is said to be *limited* in its domain.

Corresponding to a neighbourhood $(a-\epsilon, a+\epsilon)$ of a point a in a straight line, the rectangular cell which contains all points $(x_1, x_2, \dots x_n)$ such that $|x_1 - a_1| \leqq \epsilon_1$, $|x_2 - a_2| \leqq \epsilon_2$, ... $|x_n - a_n| \leqq \epsilon_n$, where $\epsilon_1, \epsilon_2, \dots \epsilon_n$ are definite positive numbers, is taken to be a neighbourhood of the point $(a_1, a_2, \dots a_n)$. A "sphere," which contains all points such that

$$(x_1 - a_1)^2 + (x_2 - a_2)^2 + \dots + (x_n - a_n)^2 \leqq \rho^2,$$

where ρ is some assigned number, is also frequently employed for purposes similar to those for which the interval $(a-\epsilon, a+\epsilon)$ is used in the case of a linear domain.

The definitions given in § 167, of the upper and lower limits, and of the fluctuation of a function in its domain, can be immediately extended to the case of an n-dimensional domain.

The function $f(x_1, x_2, \dots x_n)$ is said to be continuous at the point $(a_1, a_2, \dots a_n)$, which is a limiting point of the domain of the function, provided that, corresponding to an arbitrarily chosen positive number ϵ, a neighbourhood of $(a_1, a_2, \dots a_n)$ can be determined, such that

$$|f(x_1, x_2, \dots x_n) - f(a_1, a_2, \dots a_n)| < \epsilon,$$

for every point $(x_1, x_2, \dots x_n)$ in the interior of the neighbourhood, which is conveniently taken to be a rectangular cell.

A function which is not continuous at a point may satisfy the above condition for a neighbourhood in which $x_1 - a_1, x_2 - a_2, \dots x_n - a_n$ are restricted each to have a definite sign. The 2^n different partial neighbourhoods of a point so determined, correspond to neighbourhoods on the right and on the left, in the case of one-dimensional domains.

Such partial continuity of a function is a generalization of the conception of continuity on the right, and on the left.

The saltus of a function at a point is defined as in § 180, as the limit of the fluctuation in the neighbourhood when the greatest of the numbers $\epsilon_1, \epsilon_2, \dots \epsilon_n$ converges to zero.

There is a separate saltus for the limit of each of the 2^n partial neighbourhoods.

The domain for which a function is defined will most frequently be taken to be a continuous limited domain, *i.e.* one which is limited, perfect and connex.

The theorem of § 171, that for such a domain, there exists one point at least, such that, in any arbitrarily small neighbourhood of the point, the upper limit of the function is the same as the upper limit of the function in the whole domain, can be extended to the case of an n-dimensional domain.

The whole domain may be taken to be contained in a single rectangular cell; this rectangular cell may be sub-divided into n^r equal parts each similar to the whole; these parts may then be similarly sub-divided, and so on, indefinitely. The proof of the theorem is then precisely similar to that given in § 171.

The theorem that, in the case of a continuous function, the upper limit of the function is actually attained at some point of the domain, may be proved as in § 171.

That a continuous function is determined by the functional values at the points of an everywhere-dense enumerable set contained in its domain, may be proved as in § 173.

That a continuous function defined for a closed domain is necessarily uniformly continuous, may be proved by either of the methods employed in § 175 and § 185. Thus, if for each point of the domain, a neighbourhood be determined, within which the fluctuation of the function is less than the prescribed number ϵ, a finite number of these neighbourhoods can be selected such that each point of the domain is in the interior of one at least of them. The finite number of cells which overlap one another determine a finite number of non-overlapping cells. If η be the shortest of the edges of all these non-overlapping cells, any rectangular cell such that the lengths of all its edges are less than η will be contained in the interior of one of the cells of the finite overlapping set. Thus the theorem is established.

FUNCTIONS OF TWO VARIABLES.

230. Most of the points in which the theory of functions of a number of variables involves considerations which are not an immediate generalization of those which occur in the case of functions of a single variable, are sufficiently illustrated by the case of functions of two variables. Accordingly the properties of functions of two variables will be considered in some detail.

That a function $f(x, y)$ should be continuous at a point (a, b) which is a limiting point of its domain, it is necessary, but not sufficient, that the function $f(x, b)$ of x should be continuous at the point $x = a$, and that the function $f(a, y)$ of y should be continuous at the point $y = b$. Thus a function may be continuous at a point with respect to x, and also with respect to y, whilst it is discontinuous with respect to the two-dimensional domain (x, y).

It is not even sufficient to ensure the continuity of $f(x, y)$ at a point, that it be continuous in every direction from the point. Thus

$$f(a + r\cos\theta,\ b + r\sin\theta)$$

may be a continuous function of r, at $r = 0$, for each value of θ in the interval $(0, 2\pi)$, and yet* the function may be discontinuous at (a, b).

The necessary and sufficient condition that $f(x, y)$ may be continuous at (a, b) may be expressed in the form, that $f(x, y)$ *must be continuous in every direction at the point, and uniformly so for all directions.*

Thus if $f(a + r\cos\theta,\ b + r\sin\theta)$ be continuous at $r = 0$, for each value of θ, and uniformly so for all values of θ, then if ϵ be a prescribed positive number, a number ρ can be determined, independent of θ, such that

$$|f(a + r\cos\theta,\ b + r\sin\theta) - f(a, b)| < \epsilon,$$

provided $r < \rho$. From this condition it follows that $|f(x, y) - f(a, b)| < \epsilon$, provided $|x - a|$, $|y - b|$ are each $< \rho/\sqrt{2}$, and thus the condition of continuity of the function is satisfied.

The remarks which have been made as regards the continuity of a function at a point are applicable without essential change if those functional values in the neighbourhood of the point are alone taken into account, which are in one of the four quadrants, the values at points on the axes bounding the quadrant being either included or excluded from consideration, as may be agreed upon. Thus the condition of continuity at a point may be satisfied for one such quadrant and not for another one.

EXAMPLES.

1. Let $f(x, y) = \dfrac{xy}{x^2 + y^2}$, and $f(0, 0) = 0$. This function is discontinuous at the point $(0, 0)$, although it is continuous at that point with respect to x, and also with respect to y, since $f(x, 0) = 0$, $f(y, 0) = 0$. In all other directions the function is discontinuous; for writing $x = r\cos\theta$, $y = r\sin\theta$, the function is $\frac{1}{2}\sin 2\theta$ and therefore has a constant value different from zero on a straight line for which θ is constant, unless θ has one of the values 0, $\frac{1}{2}\pi$, π, or $\frac{3}{2}\pi$.

2. Let† $f(x, y) = \dfrac{xy^2}{x^2 + y^4}$, $f(0, 0) = 0$. This function is discontinuous at the point $(0, 0)$, although it is continuous in each particular direction, at that point. We find that $\dfrac{r\sin^2\theta}{\cos^2\theta + r^2\sin^4\theta} < \epsilon$, if $r < \dfrac{1}{2\epsilon}\operatorname{cosec}^2\theta\,\{1 - \sqrt{1 - 4\epsilon^2\cos^2\theta}\}$; and in order that this condition may be satisfied, the greatest value of r diminishes indefinitely as θ approaches the value $\frac{1}{2}\pi$; whereas when $\theta = \frac{1}{2}\pi$, the function is, for every value of r, equal to $f(0, 0)$. It is thus seen that the convergence in different directions is non-uniform.

* See Thomae, *Abriss einer Theorie der komplexen Functionen*, 2nd ed. p. 15.

† Genocchi-Peano, *Calc. Diff.*, § 128.

231. Let (a, b) be a limiting point of the domain for which a function $f(x, y)$ is defined, and let a neighbourhood of which the corners are the four points $(a \pm \epsilon, b \pm \epsilon')$ be taken. Let U, L be the upper and lower limits of the function for all points of the domain in this neighbourhood, the functional value at (a, b) being however disregarded in case (a, b) belongs to the domain. If ϵ, ϵ' be diminished, the number U cannot increase; and when values of ϵ, ϵ' belonging to sequences $\epsilon_1, \epsilon_2, \ldots \epsilon_n, \ldots$, and $\epsilon_1', \epsilon_2', \ldots \epsilon_n', \ldots$ each of which converges to the limit zero are taken, and U_n be the value of U corresponding to the values ϵ_n, ϵ_n' of ϵ, ϵ', the numbers $U_1, U_2, \ldots U_n, \ldots$ form a sequence of numbers which do not increase. This sequence has then a limit $\overline{U}$, which may however have the improper value ∞, in case all the numbers U_n have this improper value. It is easily seen that $\overline{U}$ is independent of the particular sequences chosen for ϵ, ϵ'. This number $\overline{U}$ is said to be the *upper limit of the function at* (a, b), and may be denoted by

$$\overline{\lim}_{x=a, y=b} f(x, y).$$

The lower limit $\underline{\lim}_{x=a, y=b} f(x, y)$ may be defined in a similar manner, as the limit of a sequence of values of L; and it may have the improper value $-\infty$.

At a point of continuity of the function, the condition

$$\overline{\lim}_{x=a, y=b} f(x, y) = \underline{\lim}_{x=a, y=b} f(x, y)$$

is satisfied; and further, each of these limits is equal to $f(a, b)$, in case (a, b) belongs to the domain of the function.

Corresponding pairs of limits may also be defined for the case in which the functional values in one quadrant only are taken into account, the functional values on the axes being either included or excluded, in case they exist, as may be agreed upon.

The *saltus* or *measure of discontinuity* at the point (a, b) is measured by the excess of the greatest over the least of the three numbers

$$f(a, b), \quad \overline{\lim}_{x=a, y=b} f(x, y), \quad \underline{\lim}_{x=a, y=b} f(x, y).$$

The saltus at a point of discontinuity may have a finite value, or it may be indefinitely great.

In case $\overline{\lim}_{x=a, y=b} f(x, y) = \underline{\lim}_{x=a, y=b} f(x, y)$, their common value may be denoted by $\lim_{x=a, y=b} f(x, y)$, and the function is then said to have a definite *double limit* at the point (a, b); this double limit $\lim_{x=a, y=b} f(x, y)$ may be finite, or infinite with a definite sign.

When the upper and lower limits have different values, $\lim f(x, y)$ is frequently regarded as existent but indeterminate, the upper and lower limits being regarded as its limits of indeterminacy.

232. In considering the functional values in the neighbourhood of a point, and the functional limits at the point, it is frequently convenient to consider one quadrant only; this we may take, without loss of generality, to be the quadrant in which $x-a \geqq 0$, $y-b \geqq 0$. The results which will be established are essentially applicable to any one of the four quadrants, and can be immediately extended to the case in which account is taken of the whole neighbourhood of (a, b), by taking the totality of the results for the four separate quadrants, and for the lines $x=a$, $y=b$.

Assuming that $x-a>0$, $y-b>0$, the function $f(x, y)$ considered as a function of y only, with x constant, has two functional limits $\overline{f(x, b+0)}$, $\underline{f(x, b+0)}$, at the point (x, b); these may be denoted by $\overline{\lim}_{y=b} f(x, y)$, $\underline{\lim}_{y=b} f(x, y)$ respectively. In case these two limits are identical, their common value may be denoted by $\lim_{y=b} f(x, y)$, the functional limit $f(x, b+0)$ having in that case a definite value.

If either of the limits $\overline{\lim}_{y=b} f(x, y)$, $\underline{\lim}_{y=b} f(x, y)$ is to be taken indifferently, we may denote them by $\overline{\underline{\lim}}_{y=b} f(x, y)$. This may be regarded as a function of x, such that its value at the point (x, b) is multiple-valued, and has $\overline{\lim}_{y=b} f(x, y)$, $\underline{\lim}_{y=b} f(x, y)$ for its limits of indeterminacy.

It may happen that $\overline{\underline{\lim}}_{y=b} f(x, y)$, considered as a function of x, has a definite functional limit at the point $x=a$; this limit may be either finite, or infinite with fixed sign. In case such a limit exists, it is denoted by $\lim_{x=a} \lim_{y=b} f(x, y)$, and it is said to be the *repeated limit* of $f(x, y)$ at the point (a, b), the order of the limits being, that the limit for $y=b$ is taken first, and then afterwards the limit for $x=a$.

In case this repeated limit does not exist, either as a definite number, or as infinite with fixed sign, we may regard $\lim_{x=a} \lim_{y=b} f(x, y)$ as indeterminate, its limits of indeterminacy being $\overline{\lim}_{x=a} \overline{\lim}_{y=b} f(x, y)$, and $\underline{\lim}_{x=a} \underline{\lim}_{y=b} f(x, y)$.

The repeated limit $\lim_{y=b} \lim_{x=a} f(x, y)$, in which the limit with respect to x is first taken, and afterwards that with respect to y, may be defined in a precisely similar manner.

It is clear that the functional values on the straight lines $x=a$, $y=b$ are irrelevant as regards the existence, or the values, of the repeated limits.

In case the *double limit* $\lim_{x=a,\, y=b} f(x, y)$ for $x>a$, $y>b$, exists at the point (a, b), having either a finite value, or being infinite with fixed sign, the existence of the two repeated limits

$$\lim_{x=a} \lim_{y=b} f(x, y), \quad \lim_{y=b} \lim_{x=a} f(x, y)$$

follows as a consequence, their common value being $\lim_{x=a,\, y=b} f(x, y)$. In this case $\lim_{x=a} f\{x, \phi(x)\}$ also exists, and is equal to the double limit, where $\phi(x)$ is any function of x, which is $>b$, and is such that $\lim_{x=a} \phi(x)=b$. Also $\lim_{t=\tau} f\{\phi(t), \psi(t)\}$ exists, and is equal to the double limit; where $\phi(t)$, $\psi(t)$ are functions of a variable t, such that $\phi(t)>a$, $\phi(t)>b$, and that $\lim_{t=\tau} \phi(t)=a$, $\lim_{t=\tau} \phi(t)=b$.

The converse of these statements does not hold good. In particular, the existence of $\lim_{x=a,\, y=b} f(x, y)$ is not necessary either for the existence with definite values, or for the equality, of the two repeated limits

$$\lim_{x=a} \lim_{y=b} f(x, y); \quad \lim_{y=b} \lim_{x=a} f(x, y).$$

EXAMPLES.

1. Let $f(x, y)$ be defined for the positive quadrant by $f(x, y)=\dfrac{x-y}{x+y}$. We find $\lim_{x=0} \lim_{y=0} f(x, y)=1$, $\lim_{y=0} \lim_{x=0} f(x, y)=-1$; thus $\lim_{x=0,\, y=0} f(x, y)$ cannot exist.

2. Let $f(x, y)=\dfrac{x^2y^2}{x^2y^2+(x-y)^2}$. In this case $\lim_{x=0} \lim_{y=0} f(x, y)$ and $\lim_{y=0} \lim_{x=0} f(x, y)$ are both zero, and yet $\lim_{x=0,\, y=0} f(x, y)$ does not exist; for if $y=x$, $f(x, y)=1$; and therefore $\lim_{x=0} f(x, x)=1$.

3. Let* $f(x, y)$ be defined for $x>0$, $y>0$, by the expression $(x+y)\sin\dfrac{1}{x}\sin\dfrac{1}{y}$. In this case $\overline{\lim}_{y=0} f(x, y)=x\sin\dfrac{1}{x}$, $\underline{\lim}_{y=0} f(x, y)=-x\sin\dfrac{1}{x}$, and $\overline{\lim}_{y=0} f(x, y)-\underline{\lim}_{y=0} f(x, y)$ has for $x=0$ the limit zero. We have then $\lim_{x=0} \lim_{y=0} f(x, y)=0$, since $x\sin\dfrac{1}{x}$, $-x\sin\dfrac{1}{x}$ have each the limit zero for $x=0$. It is clear that $\lim_{y=0} \lim_{x=0} f(x, y)$ is also zero. If $0<x<\frac{1}{2}\epsilon$, and $0<y<\frac{1}{2}\epsilon$, we see that $|f(x, y)|<\epsilon$, and therefore $\lim_{x=0,\, y=0} f(x, y)$ exists, and is equal to zero.

233. An important matter for investigation is the determination of the necessary and sufficient conditions for the existence and equality of the two

* Pringsheim, *Encyklopädie der Math. Wissensch.*, II A. 1, p. 51.

repeated limits at a point. A knowledge of such conditions, as also of sufficient conditions, is required in various fundamental theorems of analysis which turn upon the legitimacy of inverting the order of a repeated limiting process.

It will be observed that the existence of $\lim\limits_{x=a}\lim\limits_{y=b} f(x, y)$ does not necessarily involve the existence of $\lim\limits_{y=b} f(x, y)$ as a definite number, since $\lim\limits_{x=a}\overline{\lim\limits_{y=b}} f(x, y)$, $\lim\limits_{x=a}\underline{\lim\limits_{y=b}} f(x, y)$ may both exist and have the same value, without it being necessarily the case that $\overline{\lim\limits_{y=b}} f(x, y)$, $\underline{\lim\limits_{y=b}} f(x, y)$ are identical. It is however necessary that $\overline{\lim\limits_{y=b}} f(x, y) - \underline{\lim\limits_{y=b}} f(x, y)$ should converge to the limit zero, as x converges to the value a.

The necessary and sufficient conditions required are contained in the following general theorem:—

In order that the repeated limits $\lim\limits_{x=a}\lim\limits_{y=b} f(x, y)$, $\lim\limits_{y=b}\lim\limits_{x=a} f(x, y)$ *may both exist and have the same finite value, it is necessary and sufficient,* (1) *that* $\overline{\lim\limits_{y=b}} f(x, y) - \underline{\lim\limits_{y=b}} f(x, y)$ *should have the limit zero, for* $x = a$, *and that* $\overline{\lim\limits_{x=a}} f(x, y) - \underline{\lim\limits_{x=a}} f(x, y)$ *should have the limit zero, for* $y = b$; *and* (2) *that, corresponding to any fixed positive number* ϵ *arbitrarily chosen, a positive number* β *can be determined, such that for each value of* y *interior to the interval* $(b, b+\beta)$ *a positive number* α_y *in general dependent on* y *exists, such that, for this value of* y, $f(x, y)$ *lies between* $\overline{\lim\limits_{y=b}} f(x, y) + \epsilon$ *and* $\underline{\lim\limits_{y=b}} f(x, y) - \epsilon$, *for all values of* x *interior to the interval* $(a, a+\alpha_y)$.

Let us first assume that the conditions stated in the theorem are satisfied. A value of y may, in virtue of (1), be so chosen that the difference of the two limits $\overline{\lim\limits_{x=a}} f(x, y)$, $\underline{\lim\limits_{x=a}} f(x, y)$ is less than an arbitrarily chosen number η; and this value of y may also be so chosen that it is interior to $(b, b+\beta)$. For this fixed value of y, an interval $(a, a+\alpha_y')$ for x may be so chosen that $f(x, y)$ lies between $\overline{\lim\limits_{x=a}} f(x, y) + \epsilon$ and $\underline{\lim\limits_{x=a}} f(x, y) - \epsilon$, provided y has the fixed value, and $a < x < a + \alpha_y'$: this follows from the definition of the upper and lower limits. Again, from the condition (1), a number α'' can be determined, such that if x be interior to the interval $(a, a+\alpha'')$, the difference between the two limits $\overline{\lim\limits_{y=b}} f(x, y)$, $\underline{\lim\limits_{y=b}} f(x, y)$ is less than η. Now let $\overline{\alpha_y}$ be the smallest of the three numbers $\alpha_y, \alpha_y', \alpha''$; then, if x_1, x_2 be any two values of x within the interval $(a, a+\overline{\alpha_y})$, and y have the fixed value, by applying the conditions of the theorem, we see that the conditions

$$|f(x_1, y) - f(x_2, y)| < \eta + 2\epsilon,$$

$$\left|f(x_1, y) - \overline{\lim_{y=b}} f(x_1, y)\right| < \eta + \epsilon,$$

$$\left|f(x_2, y) - \overline{\lim_{y=b}} f(x_2, y)\right| < \eta + \epsilon,$$

are all satisfied. It follows that

$$\left|\overline{\lim_{y=b}} f(x_1, y) - \overline{\lim_{y=b}} f(x_2, y)\right| < 3\eta + 4\epsilon$$

for every pair of points x_1, x_2 within the interval $(a, a + \overline{a_y})$. Hence, since ϵ, η are both arbitrarily small, $\overline{\lim_{y=b}} f(x, y)$ converges for $x = a$ to a definite value which is the limit of both $\overline{\lim_{y=b}} f(x, y)$ and of $\underline{\lim_{y=b}} f(x, y)$ when $x = a$; and thus $\lim_{x=a} \lim_{y=b} f(x, y)$ exists.

Again, since $\lim_{x=a} \lim_{y=b} f(x, y)$ has a definite value, an interval $(a, a + \delta)$ can be determined, such that for any point x interior to it

$$\left|\lim_{x=a} \lim_{y=b} f(x, y) - \overline{\lim_{y=b}} f(x, y)\right| < \epsilon.$$

Now $\lim_{x=a} \lim_{y=b} f(x, y) - \overline{\lim_{x=a}} f(x, y)$ is the sum of the three differences

$$\lim_{x=a} \lim_{y=b} f(x, y) - \overline{\lim_{y=b}} f(x, y),$$

$$\overline{\lim_{y=b}} f(x, y) - f(x, y), \quad f(x, y) - \overline{\lim_{x=a}} f(x, y),$$

and for a fixed y, chosen as before, x may be chosen so that it not only lies within the interval $(a, a + \delta)$, but is also such that

$$\left|f(x, y) - \overline{\lim_{y=b}} f(x, y)\right|, \quad \left|f(x, y) - \overline{\lim_{x=a}} f(x, y)\right|$$

are each less than $\eta + 2\epsilon$. It follows that

$$\left|\lim_{x=a} \lim_{y=b} f(x, y) - \overline{\lim_{x=a}} f(x, y)\right| < 5\epsilon + 2\eta,$$

and thus that $\overline{\lim_{x=a}} f(x, y)$ converges, as y converges to b, to the limit $\lim_{x=a} \lim_{y=b} f(x, y)$. It has thus been shewn that the two repeated limits both exist and have the same value.

Conversely, let us assume that the repeated limits both exist, and are finite and equal. We have then $\left|\lim_{x=a} \lim_{y=b} f(x, y) - \overline{\lim_{x=a}} f(x, y)\right| < \zeta$, provided

y lies between b and $b+\beta$, where β is some fixed number, ζ being an arbitrarily chosen positive number; from this it follows that

$$\overline{\lim_{x=a}} f(x, y) - \underline{\lim_{x=a}} f(x, y)$$

is $< 2\zeta$, for $b < y < b+\beta$. Also

$$\left| \overline{\lim_{y=b}} f(x, y) - \lim_{x=a} \lim_{y=b} f(x, y) \right| < \zeta,$$

provided x lies within some fixed interval $(a, a+\delta')$; and from this it follows that $\overline{\lim\limits_{y=b}} f(x, y) - \underline{\lim\limits_{y=b}} f(x, y)$ is $< 2\zeta$, for $a < x < a+\delta'$. Since ζ is arbitrarily small, we now see that the condition (1) of the theorem is satisfied. Further we see that

$$\left| f(x, y) - \overline{\lim_{x=a}} f(x, y) \right| < 2\zeta + \zeta',$$

where ζ' is any arbitrarily chosen positive number, provided x lies within some interval $(a, a+\alpha'_y)$, where α'_y depends upon y, and may diminish indefinitely as y approaches the value b. It follows from the three inequalities, that

$$\left| f(x, y) - \overline{\lim_{y=b}} f(x, y) \right| < 4\zeta + \zeta',$$

provided $b < y < b+\beta$, and provided also x lies within some interval $(a, a+\alpha_y)$ where α_y depends in general upon y. Since ζ and ζ' are both arbitrarily small, it follows that the condition (2) of the theorem is satisfied.

If the condition (2) in the above general theorem be replaced by the more stringent condition that, corresponding to any fixed positive number ϵ, arbitrarily chosen, a positive number β can be determined, which is such that for each value of y interior to the interval $(b, b+\beta)$, a positive number α_y dependent on y exists, such that for this value of y, *and for all smaller values*, $f(x, y)$ lies between $\overline{\lim\limits_{y=b}} f(x, y) + \epsilon$, and $\underline{\lim\limits_{y=b}} f(x, y) - \epsilon$, then this condition and the condition (1) are the necessary and sufficient conditions that not only $\lim\limits_{x=a} \lim\limits_{y=b} f(x, y)$, $\lim\limits_{y=b} \lim\limits_{x=a} f(x, y)$ exist and are equal, but also that the double limit $\lim\limits_{x=a,\, y=b} f(x, y)$ exists, having a definite value the same as the repeated limits. In case the function be defined for values of x, y on the lines $x=a$, $y=b$, the additional conditions must be added that the functional values on these lines also converge to the same limit

$$\lim_{x=a,\, y=b} f(x, y).$$

For, under the conditions stated, we have, provided y lies within the interval $(b, b+\beta_1)$, where $\beta_1 < \beta$,

$$\left| f(x, y) - \overline{\lim_{y=b}} f(x, y) \right| < \epsilon + \eta,$$

where x has any value in the interval $(a, a+\zeta)$, ζ being the lesser of the two numbers α_{β_1} and δ'; the number δ' being so chosen that

$$\left|\overline{\lim_{y=b}} f(x, y) - \underline{\lim_{y=b}} f(x, y)\right| < \eta, \text{ for } a < x < a+\delta'.$$

Also $\left|\underline{\lim_{y=b}} f(x, y) - \lim_{x=a}\lim_{y=b} f(x, y)\right| < \epsilon$, provided x lies within an interval chosen sufficiently small. Hence the condition

$$\left|f(x, y) - \lim_{x=a}\lim_{y=b} f(x, y)\right| < 2\epsilon + \eta$$

is satisfied, provided $b < y < b + \beta_1$, and provided x lies within an interval of which the length may depend upon ϵ and η. It follows, since ϵ, η are arbitrarily small, that $f(x, y)$ has a definite double limit at the point (a, b). That the conditions stated are necessary, follows at once from the definition of $\lim_{x=a,\ y=b} f(x, y)$.

234. The theorem obtained in § 233 may be simplified in the case in which $\lim_{y=b} f(x, y)$, $\lim_{x=a} f(x, y)$ both have definite values at all points on the straight lines $x = a$, $y = b$ which are in sufficiently small neighbourhoods of the point (a, b). We may then state the theorem as follows:—

If $\lim_{y=b} f(x, y)$, $\lim_{x=a} f(x, y)$ *have definite finite values in the neighbourhood of the point* (a, b), *then the necessary and sufficient condition that the two repeated limits* $\lim_{x=a}\lim_{y=b} f(x, y)$, $\lim_{y=b}\lim_{x=a} f(x, y)$ *may both exist and have the same finite value is that, corresponding to any fixed positive number* ϵ, *arbitrarily chosen, a positive number* β *can be determined, which is such that, for each value of* y *interior to the interval* $(b, b+\beta)$, *a positive number* α_y *in general dependent on* y *exists, such that for this value of* y, $\left|f(x, y) - \lim_{y=b} f(x, y)\right| < \epsilon$ *for all values of* x *within the interval* $(a, a+\alpha_y)$.

In case the condition $\left|f(x, y) - \lim_{y=b} f(x, y)\right| < \epsilon$ for all values of x within $(a, a+\alpha_y)$ be satisfied not only for the particular value of y but for all smaller values, and this hold for every ϵ, then the double limit $\lim_{x=a,\ y=b} f(x, y)$ exists, and is equal to each of the repeated limits. In this case the point (a, b) is said to be *a point of uniform convergence* of the function $f(x, y)$ to the limit $\lim_{y=b} f(x, y)$ with respect to the parameter x; and thus, for such a point, there exists for each value of ϵ, an interval $(a, a+\alpha)$, where α depends in general upon ϵ, such that for each value of x within this interval, the condition $\left|f(x, y) - \lim_{y=b} f(x, y)\right| < \epsilon$, is satisfied, provided y be less than some fixed value which is the same for the whole x-interval $(a, a+\alpha)$.

It may happen that, as ϵ is indefinitely diminished, α has a positive minimum $\bar{\alpha}$. In that case the fixed interval $(a, a+\bar{\alpha})$ is such that, for each ϵ, the condition $\left|f(x, y) - \lim_{y=b} f(x, y)\right| < \epsilon$ is satisfied for all values of x within the fixed interval $(a, a+\bar{\alpha})$, provided y is less than some fixed value dependent on ϵ, the same for the whole x-interval. In this case $f(x, y)$ is said to *converge to* $\lim_{y=b} f(x, y)$ *uniformly within the interval* $(a, a+\bar{\alpha})$, with respect to the parameter x. Not only the point (a, b) but also each interior point of the interval $(a, a+\bar{\alpha})$ is then a point of uniform convergence of $f(x, y)$ to $\lim_{y=b} f(x, y)$ with respect to the parameter x.

235. The necessary and sufficient conditions for the existence and equality of the two repeated limits of $f(x, y)$ at (a, b) may be put into the following form different from that of the theorem of § 233.

The necessary and sufficient conditions that $\lim_{x=a} \lim_{y=b} f(x, y) = \lim_{y=b} \lim_{x=a} f(x, y)$, *their value being finite, are* (1) *that* $\overline{\lim_{x=a}} f(x, y)$ *converge to a definite value* $\lim_{y=b} \lim_{x=a} f(x, y)$ *when* y *converges to* b, *and that* $\overline{\lim_{y=b}} f(x, y) - \underline{\lim_{y=b}} f(x, y)$ *converge to zero, for* $x = a$; *and* (2) *that, corresponding to any arbitrarily chosen positive number* ϵ, *and to an arbitrarily chosen value* $b+\beta_0$ *of* y, *a value* $y_1 < b+\beta_0$ *of* y *can be found, and also a positive number* α, *such that the condition that* $f(x, y_1)$ *lies between*

$$\overline{\lim_{y=b}} f(x, y) + \epsilon, \text{ and } \underline{\lim_{y=b}} f(x, y) - \epsilon$$

is satisfied for every value of x *within the interval* $(a, a+\alpha)$.

In case $\lim_{y=b} f(x, y)$ *everywhere exists in the neighbourhood of* $x = a$, *the condition* (2) *is that* $\left|f(x, y_1) - \lim_{y=b} f(x, y)\right| < \epsilon$, *for every value of* x *within the interval* $(a, a+\alpha)$.

That the conditions contained in the theorem are necessary, is seen from the theorem of § 233; it will be shewn that they are sufficient. Let us assume that the conditions are satisfied. We have

$$\overline{\lim_{y=b}} f(x, y) - \lim_{y=b} \lim_{x=a} f(x, y) = \left[\overline{\lim_{y=b}} f(x, y) - f(x, y)\right]$$
$$+ \left[f(x, y) - \overline{\lim_{x=a}} f(x, y)\right] + \left[\overline{\lim_{x=a}} f(x, y) - \lim_{y=b} \lim_{x=a} f(x, y)\right].$$

A positive number β_1 can now be chosen, such that if $b < y < b+\beta_1$, the condition $\left|\overline{\lim_{x=a}} f(x, y) - \lim_{y=b} \lim_{x=a} f(x, y)\right| < \epsilon$ is satisfied; moreover we may choose β_1 so that it is $< \beta_0$.

Next, a value y_1 of y exists, such that $f(x, y_1)$ lies between $\overline{\lim}_{y=b} f(x, y) + \epsilon$, and $\underline{\lim}_{y=b} f(x, y) - \epsilon$, provided x be within the interval $(a, a+\alpha)$: the value of β_0 may be chosen so small that $\overline{\lim}_{x=a} f(x, y) - \underline{\lim}_{x=a} f(x, y) < \epsilon$, for every value of y which is $< b + \beta_0$, and therefore for the value y_1 of y. Again, an interval for x, possibly less than $(a, a + \alpha)$, can be so chosen that

$$\overline{\lim}_{y=b} f(x, y) - \underline{\lim}_{y=b} f(x, y) < \epsilon,$$

provided x lie within the interval. It follows that an interval $(a, a + \alpha')$ for x can be found, such that $\left|\overline{\lim}_{y=b} f(x, y) - f(x, y_1)\right| < 3\epsilon$. Further, the interval within which x lies may, if necessary, be so restricted that

$$\left|f(x, y_1) - \overline{\underline{\lim}}_{x=a} f(x, y_1)\right| < 2\epsilon.$$

Hence, provided x lies within a definite interval, we see that

$$\left|\overline{\underline{\lim}}_{y=b} f(x, y) - \lim_{y=b} \lim_{x=a} f(x, y)\right| < 6\epsilon;$$

and since this condition holds for an arbitrary ϵ, it follows that $\overline{\underline{\lim}}_{y=b} f(x, y)$ converges for $x = a$ to $\lim_{y=b} \lim_{x=a} f(x, y)$, and thus the sufficiency of the conditions is established.

PARTIAL DIFFERENTIAL COEFFICIENTS.

236. If, at a point (x_0, y_0) in the domain for which the function $f(x, y)$ is defined, the limit $\lim_{h=0} \dfrac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}$ exists, having either a definite finite value, or being indefinitely great but of fixed sign, this limit is said to be the *partial differential coefficient* of $f(x, y)$ at (x_0, y_0) with respect to x, and is usually denoted by $\dfrac{\partial f(x_0, y_0)}{\partial x_0}$.

When the limit $\lim_{k=0} \dfrac{f(x_0, y_0 + k) - f(x_0, y_0)}{k}$ exists, it is said to be the partial differential coefficient of $f(x, y)$ at (x_0, y_0) with respect to y, and is denoted by $\dfrac{\partial f(x_0, y_0)}{\partial y_0}$.

In general, h, k in these definitions are regarded as having either sign. It is possible that either of the above limits may not exist, but that there may be two definite limits, one for positive values of the increment h or k, and the other for negative values. In that case the two limits are said to be the progressive and regressive partial differential coefficients with respect to the particular variable. It is of course possible that, at a particular point, one of these may exist, and not the other.

That the two partial differential coefficients $\frac{\partial f}{\partial x_0}$, $\frac{\partial f}{\partial y_0}$ may exist, it is necessary, but not sufficient, that $f(x, y)$ should, at the point (x_0, y_0), be continuous with respect to x, and also with respect to y.

To express the increment $f(x_0+h, y_0+k) - f(x_0, y_0)$ of the function $f(x, y)$, when the two numbers x_0, y_0 receive increments h, k respectively, we have

$$f(x_0+h, y_0+k) - f(x_0, y_0) = [f(x_0+h, y_0+k) - f(x_0, y_0+k)] + [f(x_0, y_0+k) - f(x_0, y_0)].$$

If we now assume that $\frac{\partial f}{\partial y}$ exists at the point (x_0, y_0), and has a finite value, we have

$$\frac{f(x_0, y_0+k) - f(x_0, y_0)}{k} = \frac{\partial f(x_0, y_0)}{\partial y_0} + \sigma(k),$$

where $\sigma(k)$ converges to the limit zero, when k is indefinitely diminished.

Again, $\frac{f(x_0+h, y_0+k) - f(x_0, y_0+k)}{h}$ converges to the limit $\frac{\partial f}{\partial x_0}$, when k is first diminished to the limit zero, and afterwards h converges to zero, it being assumed that $\frac{\partial f}{\partial x_0}$ has a definite value, and also that $f(x, y)$ is continuous with respect to y for the value $y = y_0$, where x has any value in a neighbourhood of x_0. In order, however, that the double limit

$$\lim_{h=0,\, k=0} \frac{f(x_0+h, y_0+k) - f(x_0, y_0+k)}{h}$$

may exist, in which case its value is $\frac{\partial f}{\partial x_0}$, being independent of the mode in which h, k approach their limits, it is necessary and sufficient that

$$\frac{f(x_0+h, y_0+k) - f(x_0, y_0+k)}{h}$$

should be a continuous function of (h, k) at the point $h=0$, $k=0$.

If this condition be satisfied, positive numbers h_1, k_1 can be determined, such that

$$\left| \frac{f(x_0+h, y_0+k) - f(x_0, y_0+k)}{h} - \frac{\partial f}{\partial x_0} \right| < \eta,$$

where η is a prescribed positive number, and $0 < |h| \leqq h_1$, $0 \leqq |k| \leqq k_1$.

We have now

$$\frac{f(x_0+h, y_0+k) - f(x_0, y_0+k)}{h} = \frac{\partial f}{\partial x_0} + \rho(h, k),$$

where $\rho(h, k)$ converges to zero, independently of the mode in which h, k converge to zero.

Under the conditions stated, we have

$$f(x_0+h, y_0+k)-f(x_0, y_0)=h\frac{\partial f}{\partial x_0}+k\frac{\partial f}{\partial y_0}+h\rho+k\sigma$$

where ρ, σ converge to zero, when h and k are indefinitely diminished, independently of the mode in which they approach their limits. This is equivalent to the statement that, corresponding to an arbitrarily assigned positive number η, positive numbers h_1, k_1 can be determined so that $|\rho|$ and $|\sigma|$ are each $<\eta$, for all values of h and k such that $|h|<h_1$, $|k|<k_1$.

In the notation of differentials, denoting $f(x_0, y_0)$ by z_0, we have

$$dz_0=\frac{\partial z}{\partial x_0}dx+\frac{\partial z}{\partial y_0}dy;$$

the expression on the right-hand side being termed the *total differential* of z at the point (x, y), and $\frac{\partial z}{\partial x_0}dx$, $\frac{\partial z}{\partial y_0}dy$ the *partial differentials*. In accordance with the arithmetical theory, this equation can only be regarded as a conveniently abridged form of the result obtained in the present discussion.

The theorem obtained may be stated as follows:—

The increment of a function $f(x_0, y_0)$ when x_0, y_0 are changed into x_0+h, y_0+k is $h\frac{\partial f(x_0, y_0)}{\partial x_0}+k\frac{\partial f(x_0, y_0)}{\partial y_0}+h\rho+k\sigma$, where ρ, σ converge to zero when h, k are indefinitely diminished, independently of the mode in which they are diminished, provided that (1) $\frac{\partial f(x_0, y_0)}{\partial x_0}$, $\frac{\partial f(x_0, y_0)}{\partial y_0}$ *have definite finite values, and* (2) *that* $\frac{f(x_0+h, y_0+k)-f(x_0, y_0+k)}{h}$ *is a continuous function of* (h, k) *at the point* $h=0$, $k=0$.

It will be observed that no assumption has been made that $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$ have definite values except at the point (x_0, y_0) itself.

If it be assumed that $\frac{\partial f}{\partial x}$ has a definite value at (x_0, y) for all values of y in some neighbourhood of y_0, the condition (2) may be expressed in the form that (a) $\frac{\partial f}{\partial x}$ for $x=x_0$ must be a continuous function of y at $y=y_0$, and that (b) the point $h=0$, $k=0$ must be a point of uniform convergence of the function $\frac{f(x_0+h, y_0+k)-f(x_0, y_0+k)}{h}$ considered as a function of h, to its limit for $h=0$, with k as a parameter, in accordance with the definition of such a point of uniform convergence given in § 234.

If it be assumed that $\frac{\partial f}{\partial x}$ exists, not merely at the point (x_0, y_0), but at all points in a sufficiently small two-dimensional neighbourhood of the point, the conditions contained in the theorem may be simplified. For we have, in that case, $\frac{f(x_0+h, y_0+k)-f(x_0, y_0+k)}{h}=\frac{\partial}{\partial x}f(x_0+\theta h, y_0+k)$,

where θ is such that $0<\theta<1$; and this expression converges to $\frac{\partial f(x_0, y_0)}{\partial x_0}$, provided $\frac{\partial f(x, y)}{\partial x}$ be continuous with respect to (x, y), at the point (x_0, y_0).

It has thus been proved that*, *in order that the increment of the function may be of the form given in the theorem above, it is sufficient that* (1) $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$ *have definite values at the point* (x_0, y_0), *and* (2) *that one at least of these partial differential coefficients have definite values everywhere in a two-dimensional neighbourhood of* (x_0, y_0), *and be continuous at* (x_0, y_0) *with respect to the domain* (x, y).

237. Let it now be assumed that, throughout a perfect and connex domain D, the two partial differential coefficients $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$ everywhere exist, and that they are continuous functions of (x, y). We have then

$$f(x+h, y+k)-f(x, y+k)=h\frac{\partial}{\partial x}f(x+\theta h, y+k),$$

$$f(x, y+k)-f(x, y)=k\frac{\partial}{\partial y}f(x, y+\theta_1 k),$$

when θ, θ_1 are proper fractions, provided (x, y) is a point of D, and h, k are so chosen that the straight line joining $(x, y+k)$, $(x+h, y+k)$, and the straight line joining (x, y), $(x, y+k)$ are wholly in the domain D. Since

$$\frac{\partial}{\partial x}f(x, y),\ \frac{\partial}{\partial y}f(x, y)$$

are continuous functions of (x, y), it follows that they are uniformly continuous in the domain D. A positive number δ can accordingly be determined, corresponding to a prescribed positive number ϵ, so that

$$\left|\frac{\partial}{\partial x}f(x+\theta h, y+k)-\frac{\partial}{\partial x}f(x, y)\right|<\epsilon,$$

$$\left|\frac{\partial}{\partial y}f(x, y+\theta_1 k)-\frac{\partial}{\partial y}f(x, y)\right|<\epsilon,$$

* Thomae, *Einleitung in die Theorie der bestimmten Integrale*, p. 37.

provided $|h|$, $|k|$ are each less than δ, whatever be the position of (x, y) in D.

We thus find that

$$f(x+h,\ y+k)-f(x,\ y)=h\frac{\partial}{\partial x}f(x,\ y)+k\frac{\partial}{\partial y}f(x,\ y)+hR+kR',$$

where R and R' tend to the limit zero, with h and k, uniformly for all points (x, y) of the domain D.

This equation holds for every point (x, y) of D, and for all values of h and k, such that the straight line joining

$$(x,\ y),\quad (x,\ y+k)$$

and the straight line joining $(x, y+k)$, $(x+h, y+k)$ lie wholly in the domain D. It is easy to replace the last condition by the less stringent one that the two points $P(x, y)$, $Q(x+h, y+k)$, of the domain D can be joined by a number of straight lines PP_1, P_1P_2, P_2P_3, ... P_nQ, each of which is parallel to one of the axes, all of which belong to D, and are such that all of them are wholly interior to a rectangle with its corners at P and Q, and sides parallel to the axes.

We have then

$$f(P_1)-f(P)=h_1\frac{\partial}{\partial x}f(x+\theta_1 h_1,\ y),$$

$$f(P_2)-f(P_1)=k_1\frac{\partial}{\partial y}f(x+h_1,\ y+\theta_2 k_1),$$

$$f(P_3)-f(P_2)=h_2\frac{\partial}{\partial x}f(x+h_1+\theta_3 h_2,\ y+k_1),$$

....

where h_1, k_1 ... are the lengths of PP_1, P_1P_2, ..., and θ_1, θ_2, ... are proper fractions. Now $|h_1|$, $|k_1|$, $|h_2|$ ... all being less than δ, all the partial differential coefficients on the right-hand side of these equations differ numerically from the corresponding partial differential coefficient

$$\frac{\partial}{\partial x}f(x,\ y),\ \text{ or }\ \frac{\partial}{\partial y}f(x,\ y)$$

by less than ϵ. We thus see that

$$f(x+h,\ y+k)-f(x,\ y)=h\frac{\partial}{\partial x}f(x,\ y)+k\frac{\partial}{\partial y}f(x,\ y)+hR+kR',$$

where $|R|$, $|R'|$ tend to the limit zero, with h and k, uniformly for all points of D.

EXAMPLES.

1. Let* $f(x, y)=\sqrt{|xy|}$, where the positive value of the square root is to be taken. In this case $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$ both exist at the point (0, 0) and are both $=0$. We have

$$\frac{f(h, k)-f(0, k)}{h}=\sqrt{\left|\frac{k}{h}\right|};$$

and this has different constant values for different constant values of k/h, and is therefore discontinuous at the point $h=0$, $k=0$. It follows that the equation $f(h, k)=h\rho+k\sigma$, when ρ and σ converge to zero with h and k, cannot hold.

2. Let† $f(x, y)=x\sin(4\tan^{-1}y/x)$, for $x>0$; and $f(0, y)=0$, for all values of y. We find $\frac{\partial f(0, 0)}{\partial x}=0$, $\frac{\partial f(0, y)}{\partial x}=0$, and thus $\frac{\partial f(0, y)}{\partial x}$ is continuous with respect to y at (0, 0). Also, we find $\frac{\partial f(x, 0)}{\partial y}=4$, $\frac{\partial f(0, 0)}{\partial y}=0$, and therefore $\frac{\partial f(x, 0)}{\partial y}$ is discontinuous with regard to x, at (0, 0). The value of $\frac{f(h, k)-f(0, k)}{h}$ is $\sin\left(4\tan^{-1}\frac{k}{h}\right)$, and this is discontinuous at $h=0$, $k=0$; hence the relation $df=\frac{\partial f}{\partial x}dx+\frac{\partial f}{\partial y}dy$, does not hold at the point (0, 0).

HIGHER PARTIAL DIFFERENTIAL COEFFICIENTS.

238. If the function $f(x, y)$ have the partial differential coefficient $\frac{\partial f}{\partial x}$, it may happen that, at the point (x_0, y_0), the function $\frac{\partial f}{\partial x}$ possesses a differential coefficient with respect to x. This is denoted by $\frac{\partial^2 f(x_0, y_0)}{\partial x_0^2}$, and is spoken of as the *second partial differential coefficient* of $f(x, y)$ with respect to x, at the point x_0. The second partial differential coefficient $\frac{\partial^2 f(x_0, y_0)}{\partial y_0^2}$ with respect to y, is defined in a similar manner.

It may happen that, at the point (x_0, y_0), the function $\frac{\partial f}{\partial x}$ has a differential coefficient with respect to y: this may be denoted by $\frac{\partial}{\partial y_0}\left(\frac{\partial f}{\partial x_0}\right)$ or $\frac{\partial^2 f(x_0, y_0)}{\partial y_0 \partial x_0}$. Similarly, when $\frac{\partial f}{\partial y}$ has, at the point (x_0, y_0), a partial differential coefficient with respect to x, this is denoted by $\frac{\partial}{\partial x_0}\left(\frac{\partial f}{\partial y_0}\right)$ or $\frac{\partial^2 f(x_0, y_0)}{\partial x_0 \partial y_0}$. These partial differential coefficients are said to be the *mixed partial differential coefficients* of the second order at (x_0, y_0) of $f(x, y)$ with respect to x and y, the order of differentiation being different in the two.

* Stolz, *Grundzüge*, vol. I. p. 133.

† Harnack's *Introduction to the Differential and Integral Calculus*, Cathcart's Translation, p. 93.

Under certain conditions which will be here investigated, the two mixed partial differential coefficients of the second order with respect to x and y satisfy the relation

$$\frac{\partial^2 f(x, y)}{\partial x\, \partial y} = \frac{\partial^2 f(x, y)}{\partial y\, \partial x},$$

which is known as the fundamental theorem for partial differential coefficients of the second order.

The differential coefficient $\frac{\partial}{\partial x_0}\left(\frac{\partial f}{\partial y_0}\right)$, or $\frac{\partial^2 f(x_0, y_0)}{\partial x_0\, \partial y_0}$, is the repeated limit

$$\lim_{h=0}\lim_{k=0}\frac{f(x_0+h, y_0+k) - f(x_0+h, y_0) - f(x_0, y_0+k) + f(x_0, y_0)}{hk}.$$

We may denote this repeated limit by $\lim\limits_{h=0}\lim\limits_{k=0} F(h, k)$. In order that the partial differential coefficient may exist, the value of this limit must be independent of the signs of h and k.

It should be observed that it is not essential for the convergence of this repeated limit to a definite finite value, that

$$\lim_{k=0}\frac{f(x_0+h, y_0+k) - f(x_0+h, y_0)}{k}$$

should have a definite value when $h \neq 0$. Thus the repeated limit may have a definite value when

$$\lim_{h=0}\frac{1}{h}\left[\overline{\lim_{k=0}}\frac{f(x_0+h, y_0+k) - f(x_0+h, y_0)}{k} - \underline{\lim_{k=0}}\frac{f(x_0+h, y_0+k) - f(x_0+h, y_0)}{k}\right]$$

vanishes. The repeated limit cannot however have a definite finite value unless $\lim\limits_{k=0}\frac{f(x_0, y_0+k) - f(x_0, y_0)}{k}$ has a definite finite value, *i.e.* unless $\frac{\partial f(x_0, y_0)}{\partial y_0}$ exists and is finite. It thus appears that $\frac{\partial^2 f}{\partial x_0\, \partial y_0}$ may exist when $\frac{\partial f}{\partial y}$ exists at the point (x_0, y_0), but is indefinite at points in the neighbourhood. The existence of the repeated limit as a definite number implies the existence of $\frac{\partial f}{\partial y_0}$ and of $\frac{\partial^2 f}{\partial x_0\, \partial y_0}$.

If, however, the repeated limit $\lim\limits_{h=0}\lim\limits_{k=0} F(h, k)$ be infinite, with a definite sign, we cannot infer that $\frac{\partial^2 f}{\partial x_0\, \partial y_0}$ exists, with an infinite value, unless it be postulated that $\frac{\partial f}{\partial y}$ has a definite value at (x_0, y_0); for the existence of $\frac{\partial f}{\partial y}$ at the point cannot, in this case, be inferred from that of the repeated limit; and unless $\frac{\partial f}{\partial y}$ exists at the point, $\frac{\partial^2 f}{\partial x_0 \partial y_0}$ has not been defined. When this condition is satisfied, the value of $\frac{\partial^2 f}{\partial x_0 \partial y_0}$ is infinite with definite sign.

239. The differential coefficient $\frac{\partial^2 f}{\partial y_0 \partial x_0}$ is the repeated limit

$$\lim_{k=0} \lim_{h=0} \frac{f(x_0+h, y_0+k) - f(x_0+h, y) - f(x_0, y_0+k) + f(x_0, y_0)}{hk};$$

and thus the conditions that the relation $\frac{\partial^2 f}{\partial x_0 \partial y_0} = \frac{\partial^2 f}{\partial y_0 \partial x_0}$ holds are identical with the conditions that the two repeated limits may be identical. The necessary and sufficient conditions may be accordingly obtained by applying the conditions contained in either of the theorems in § 233, and § 235, to the function $F(h, k) \equiv \frac{f(x_0+h, y_0+k) - f(x_0+h, y) - f(x_0, y_0+k) + f(x_0, y_0)}{hk}$.

It is however convenient, for application in particular cases, to have sufficient conditions relating to the partial differential coefficients in the neighbourhood of the point (x_0, y_0).

The following theorem will be established:—

If (1), $\frac{\partial^2 f(x, y)}{\partial y \partial x}$ *exist and be finite at all points in a two-dimensional neighbourhood of the point* (x_0, y_0), *except that its existence at* (x_0, y_0) *is not assumed, and* (2) *the point* (x_0, y_0) *be a point of continuity of* $\frac{\partial^2 f(x, y)}{\partial y \partial x}$ *with respect to* (x, y), *the limit of this partial differential coefficient at* (x_0, y_0) *being a definite number* A, *and* (3) $f(x, y)$ *be continuous with respect to* x *at* (x_0, y_0), *then* $\frac{\partial^2 f(x_0, y_0)}{\partial y_0 \partial x_0}$, $\frac{\partial^2 f(x_0, y_0)}{\partial x_0 \partial y_0}$ *both exist, and have the same value* A.

It will be observed that the condition (1) implies the existence of $\frac{\partial f(x, y)}{\partial x}$ at all points in a neighbourhood of (x_0, y_0), except at that point itself, and that it is continuous with respect to y.

From the condition (2), we have, corresponding to an arbitrarily chosen positive number ϵ,

$$\frac{\partial^2 f(x_0+h, y_0+k)}{\partial y \partial x} = A + \alpha(h, k);$$

where $|\alpha| < \epsilon$, provided $|h|$, $|k|$ are each less than some fixed positive number η dependent upon ϵ, and are not both zero.

Let $u(k')$ denote $\frac{\partial f(x_0+h, y_0+k')}{\partial x} - Ak'$, where k' lies in the interval

$(0, k)$; we have then $\frac{du(k')}{dk'} = \alpha(h, k')$, and this is numerically $< \epsilon$. It follows that $\frac{u(k) - u(0)}{k}$ is numerically $< \epsilon$; for, by the mean value theorem of § 203, since $u(k')$ is continuous at $k' = 0$, and at $k' = k$, and possesses a definite differential coefficient at every interior point of the interval $(0, k)$, there exists a number $\bar{k}$ in the interval $(0, k)$ such that $\frac{u(k) - u(0)}{k} = \frac{du(\bar{k})}{dk}$, and this is numerically less than ϵ. We have now

$$\frac{\partial f(x_0 + h, y_0 + k)}{\partial x} - \frac{\partial f(x_0 + h, y_0)}{\partial x} - Ak = k\alpha''(h, k),$$

where α'' is numerically $< \epsilon$. This holds for each value of h such that $0 < |h| < \eta$.

Let $v(h')$ denote $\frac{f(x_0 + h', y_0 + k) - f(x_0 + h', y_0)}{k} - Ah'$, where h' lies in the interval $(0, h)$; we have then $\frac{dv(h')}{dh'} = \alpha''(h', k)$, and this is numerically less than ϵ. As before, since $v(h')$ is, in virtue of (3), continuous at $h' = 0$, and also at $h' = h$, and possesses a definite differential coefficient at all interior points of the interval $(0, h)$, it follows that $\frac{v(h) - v(0)}{h}$ is numerically $< \epsilon$; hence

$$hkF(h, k) = f(x_0 + h, y_0 + k) - f(x_0 + h, y_0) - f(x_0, y_0 + k) + f(x_0, y_0)$$
$$= Ahk + hk\alpha'''(h, k),$$

where α''' is numerically less than ϵ.

We have now, corresponding to the arbitrarily chosen ϵ, $|F(h, k) - A| < \epsilon$, provided h, k are each numerically less than some fixed number η dependent on ϵ. It follows that $F(h, k)$ is continuous at the point $h = 0$, $k = 0$ in the two-dimensional domain (h, k), and has A for its double limit. From this we conclude that the two limits $\lim_{h=0} \lim_{k=0} F(h, k)$, $\lim_{k=0} \lim_{h=0} F(h, k)$ exist, and are both identical with A. It follows that, when the conditions stated in the theorem are satisfied, the two partial differential coefficients $\frac{\partial^2 f(x_0, y_0)}{\partial x_0 \partial y_0}$, $\frac{\partial^2 f(x_0, y_0)}{\partial y_0 \partial x_0}$ both exist and are equal to A. The existence of $\frac{\partial f}{\partial x_0}$, $\frac{\partial f}{\partial y_0}$ follows from the existence of the above partial differential coefficients at the point (x_0, y_0).

The sufficient conditions in the foregoing theorem are somewhat simpler than those stated by Schwarz*, who assumed the additional condition that $\frac{\partial f(x, y_0)}{\partial y_0}$ exists and is finite for values of x in the neighbourhood of $x = x_0$, for

* *Gesammelte Abh.*, vol. II. p. 275; see also Peano, *Mathesis*, vol. X. p. 153. See further Stolz, *Grundzüge d. Diff. Rech.* vol. I. p. 147.

the constant value y_0. Schwarz's theorem is, however, more general, in that it is applicable to the case in which the two partial differential coefficients have an infinite value with definite sign. The method of the above proof may, however, be extended to this case, as follows:—

Let us assume that, if M be an arbitrarily chosen positive number, the condition $\frac{\partial^2 f(x_0+h, y_0+k)}{\partial y\, \partial x} > M$, is satisfied for all values of h and k which are not both zero, and are both numerically less than some fixed number η, dependent on M. Defining $u(k')$ as $\frac{\partial f(x_0+h, y_0+k')}{\partial x}$, we see, by means of the mean value theorem, that

$$\frac{u(k)-u(0)}{k} > M, \text{ or } \frac{\partial f(x_0+h, y_0+k)}{\partial x} - \frac{\partial f(x_0+h, y_0)}{\partial x} > kM.$$

Next, defining $v(h')$ as $\frac{f(x_0+h', y_0+k) - f(x_0+h', y_0)}{k}$, we see, as before, that $\frac{v(h)-v(0)}{h} > M$; therefore $F(h, k) > M$, provided h, k are both numerically less than some fixed number η dependent on M. It follows that $F(h, k)$ converges to the limit $+\infty$, with fixed sign, as h, k converge in any manner, each to the limit zero; thus both the limits $\lim_{h=0} \lim_{k=0} F(h, k)$, $\lim_{k=0} \lim_{h=0} F(h, k)$ are $+\infty$. In order that $\frac{\partial^2 f(x_0, y_0)}{\partial x_0\, \partial y_0}$, $\frac{\partial^2 f(x_0, y_0)}{\partial y_0\, \partial x_0}$ may exist, in which case they both have the value $+\infty$, it is necessary to assume that $\frac{\partial f(x_0, y_0)}{\partial x_0}$, $\frac{\partial f(x_0, y_0)}{\partial y_0}$ both have definite values. The case in which the limits are both $-\infty$ may be treated in a precisely similar manner. The following theorem has now been established:—

If (1) $\frac{\partial^2 f(x, y)}{\partial y\, \partial x}$ *exist and be finite at all points in a two-dimensional neighbourhood of the point* (x_0, y_0), *except that its existence at* (x_0, y_0) *is not assumed, and* (2) *the function* $\frac{\partial^2 f(x, y)}{\partial y\, \partial x}$ *have the limit* $+\infty$ *or* $-\infty$, *with definite sign, at the point* (x_0, y_0), *and* (3) *the differential coefficients* $\frac{\partial f(x_0, y_0)}{\partial x_0}$, $\frac{\partial f(x_0, y_0)}{\partial y_0}$ *both exist and have definite values, then* $\frac{\partial^2 f(x_0, y_0)}{\partial x_0\, \partial y_0}$, $\frac{\partial^2 f(x_0, y_0)}{\partial y_0\, \partial x_0}$ *both exist, having the value* $+\infty$, *or* $-\infty$, *with definite sign.*

240. The partial differential coefficients of higher order n of a function $f(x, y)$ are of the form $\frac{\partial^n f(x, y)}{\partial x^p\, \partial y^q\, \partial x^r \ldots \partial x^k\, \partial y^l}$, where $p, q, r, \ldots l$ are positive integers, including zero, such that $p+q+r+\ldots+l=n$. Here, f is first differentiated l times with respect to y, then k times with respect to x, and so on. The total number of possible partial differential coefficients of order n is 2^n; the number of those in which r differentiations with respect to x, and $n-r$ with respect to y are involved is $\dfrac{n!}{r!\,(n-r)!}$.

Sufficient conditions for the existence of all the partial coefficients of order n may be obtained by extending the theorem of § 239, which refers to the case $n=2$. The following criteria* which can be proved by induction, will be sufficient for the purpose:—

If the $n-1$ differential coefficients $\dfrac{\partial^n f}{\partial x^{n-1}\partial y}, \dfrac{\partial^n f}{\partial x^{n-2}\partial y^2}, \ldots \dfrac{\partial^n f}{\partial x\,\partial y^{n-1}}$ have definite finite values for all points in a two-dimensional neighbourhood of the point (x_0, y_0), and are continuous at the point (x_0, y_0) with respect to (x, y), then all the other mixed partial differential coefficients of order n exist at the point (x_0, y_0); and each one of them has the same value at the point as that one of those given above in which the same number of differentiations with respect to x, and with respect to y, occurs, as in the one considered.

EXAMPLE.

Let† the function $f(x, y)$ be defined by $f(x, y)=xy\dfrac{x^2-y^2}{x^2+y^2}$, for all values of x and y except when $x=0, y=0$; for which $f(0, 0)=0$. At the point $(0, 0)$, the partial differential coefficients $\dfrac{\partial^2 f}{\partial x\partial y}, \dfrac{\partial^2 f}{\partial y\partial x}$ both exist, and have different finite values.

The function $f(x, y)$ is continuous at the point $(0, 0)$; for, writing $x=r\cos\theta, y=r\sin\theta$, the function becomes $\frac{1}{4}r^2\sin 4\theta$; and this is numerically less than ϵ, provided $r<2\sqrt{\epsilon}$.

We find $\dfrac{\partial f(x, y)}{\partial x}=y\left\{\dfrac{x^2-y^2}{x^2+y^2}+\dfrac{4x^2y^2}{(x^2+y^2)^2}\right\}$, at any point except $(0, 0)$; at which point $\dfrac{\partial f}{\partial x}$ is $\lim\limits_{x=0}\dfrac{f(x, 0)-f(0, 0)}{x}$, which is $=0$.

The value of $\dfrac{\partial f(0, y)}{\partial x}$ is $-y$, and that of $\dfrac{\partial f(x, 0)}{\partial y}$ is x.

We then find
$$\frac{\partial^2 f(0, 0)}{\partial y\partial x}=\lim_{y=0}\frac{1}{y}\left\{\frac{\partial f(0, y)}{\partial x}-\frac{\partial f(0, 0)}{\partial x}\right\}=-1,$$
and
$$\frac{\partial^2 f(0, 0)}{\partial x\partial y}=\lim_{x=0}\frac{1}{x}\left\{\frac{\partial f(x, 0)}{\partial y}-\frac{\partial f(0, 0)}{\partial y}\right\}=1.$$

The value of $\dfrac{\partial^2 f(x, y)}{\partial x\partial y}$, as also that of $\dfrac{\partial^2 f(x, y)}{\partial y\partial x}$ is $\dfrac{x^2-y^2}{x^2+y^2}\left\{1+\dfrac{8x^2y^2}{(x^2+y^2)^2}\right\}$, at every point except $(0, 0)$. This value is $\cos 2\theta\,(1+2\sin^2 2\theta)$, which is constant for a constant value of θ, but has different values for different values of θ; and thus the partial differential coefficients are discontinuous at the point $(0, 0)$. The conditions of the theorem giving sufficient conditions for the equality of $\dfrac{\partial^2 f}{\partial x\partial y}$ and $\dfrac{\partial^2 f}{\partial y\partial x}$ are therefore not satisfied for the point $(0, 0)$.

* See Stolz, *Grundzüge*, vol. I., p. 153.

† Peano, *Calc. Diff.*, p. 174.

MAXIMA AND MINIMA OF A FUNCTION OF TWO VARIABLES.

241. Let us suppose that a function $f(x, y)$ is defined at all points in a two-dimensional neighbourhood of the point (x_0, y_0).

If the function be such that $f(x_0+h, y_0+k)-f(x_0, y_0)<0$, for all values of h, k which are not both zero, and are such that $|h|$, $|k|$ are both less than some fixed positive number δ, then the function $f(x, y)$ is said to have *a proper maximum* at the point (x_0, y_0).

In case the fixed number δ can only be so determined that the condition $f(x_0+h, y_0+k)-f(x_0, y_0) \leqq 0$, is satisfied, the function is said to have *an improper maximum* at the point (x_0, y_0).

If the conditions contained in these definitions be replaced by

$$f(x_0+h, y_0+k)-f(x_0, y_0)>0, \text{ and } f(x_0+h, y_0+k)-f(x_0, y_0) \geqq 0$$

respectively, the function $f(x, y)$ is said to have, in the first case, *a proper minimum*, and in the second case, *an improper minimum*, at the point (x_0, y_0).

A proper or improper maximum or minimum may be spoken of as an *extreme* of the function.

At an extreme (x_0, y_0), $f(x_0+h, y_0)-f(x_0, y_0)$, $f(x_0-h, y_0)-f(x_0, y_0)$ both have the same sign, or are zero, for all sufficiently small values of h; it follows that, if $\dfrac{\partial f(x_0, y_0)}{\partial x_0}$ exist, it must be zero. A similar remark applies to $\dfrac{\partial f(x_0, y_0)}{\partial y_0}$.

These conditions are necessary, under the hypothesis of the existence of the two partial differential coefficients, but not sufficient, for the existence of an extreme at the point (x_0, y_0).

If we write $x=x_0+r\cos\theta$, $y=y_0+r\sin\theta$, $f(x, y)=\phi(r, \theta)$, it is clearly necessary for the existence of an extreme of $f(x, y)$ at (x_0, y_0), that $\phi(r, \theta)$ for each constant value of θ, should have an extreme at $r=0$. Thus, for an assigned value of θ, a positive number a_θ can be determined such that one of the four conditions $\phi(r, \theta)-f(x_0, y_0)<0$, $\phi(r, \theta)-f(x_0, y_0) \leqq 0$, $\phi(r, \theta)-f(x_0, y_0)>0$, $\phi(r, \theta)-f(x_0, y_0) \geqq 0$, according as the point is a proper maximum, an improper maximum, a proper minimum, or an improper minimum, shall be satisfied for all values of r different from zero, and such that $|r|<a_\theta$. Thus an extreme of a function is necessarily an extreme for values of the function on each straight line drawn through the point.

This condition, though necessary, is however not sufficient; for a_θ may have a definite value for each value of θ, and yet the lower limit of a_θ for all values of θ may be zero. In this case, no value of δ can be determined, as required in the definition of the extreme in the two-dimensional domain. It has thus been shewn that, *in order that* (x_0, y_0) *may be an extreme point* *for the function $f(x, y)$, it is necessary and sufficient that* (1) $r = 0$ *should be an extreme point of $\phi(r, \theta)$ for each value of θ, and* (2) *that* the number a_θ which is so determined for each value of θ that for $|r| < a_\theta$ the condition as to $\phi(r, \theta) - f(x_0, y_0)$ may be satisfied, should have a finite lower limit when all values of θ, $(0 \leqq \theta \leqq \pi)$ are considered. If the lower limit of a_θ be zero, the point is not an extreme point of the function.*

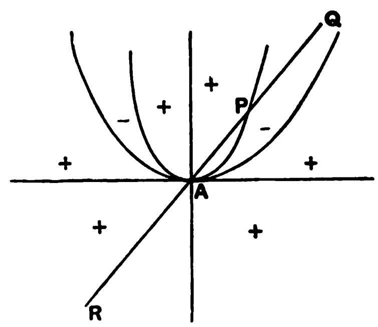
When the lower limit of a_θ is $d (> 0)$, the neighbourhood of (x_0, y_0), which must exist in accordance with the definition, is the square of which the corners are the four points $\left(x_0 \pm \frac{d}{\sqrt{2}}, y_0 \pm \frac{d}{\sqrt{2}}\right)$.

EXAMPLE.

As an example of a function which possesses no minimum at a point, although the point is a minimum for each straight line through the point, we may take the function

$$(y - ax^2)(y - bx^2) \equiv y^2 - y(ax^2 + bx^2) + abx^4,$$

where a and b have positive values.



The function is positive outside the two parabolas

$$y - ax^2 = 0, \quad y - bx^2 = 0,$$

and in the space interior to the inner parabola; in the space between the parabolas, the function is negative. Along any straight line QAR through A (0, 0), the function exceeds $f(0, 0)$ at all points interior to AP, and everywhere in PA produced; thus for the line QAR the function has a minimum at A. The point (0, 0) is not a minimum of the function, since the lower limit of AP for all positions of QAR is zero; and thus there exists no two-dimensional neighbourhood of A, in which the function is never less than at A.

242. We may without loss of generality take the point at which the conditions for the existence of an extreme of the function $f(x, y)$ are to be investigated as the point (0, 0). It will be assumed that, at all points in the neighbourhood of (0, 0), $f(x, y)$ is continuous with respect to x, and also with respect to y. The following theorem contains a criterion for the existence of a proper maximum (minimum) at the point (0, 0).

The necessary and sufficient conditions that the point (0, 0) may be a point at which $f(x, y)$ has a proper maximum (minimum) are the following†:—(1) *A positive number δ must exist which is such that, if x be any*

* The necessity for this condition has been disregarded in many text-books. The insufficiency of (1) was first pointed out by Peano, *Calcolo diff.*, Turin 1884, p. 29, in connection with the example given in the text. See also Dantscher, *Math. Annalen*, vol. XLII., p. 89, and Scheeffer, *Math. Annalen*, vol. XXXV., p. 541.

† See Stolz, *Wiener Berichte* (Nachtrag), vol. 100, also *Grundzüge*, vol. I., p. 213.

number different from zero, and numerically less than δ, *the upper (lower) limit of* $f(x, y)$, *for such constant value of* x, *and for all values of* y *for which* $-x \leqq y \leqq x$, *being* $f(x, \phi(x))$, *this upper (lower) limit is for every value of* $x\ (-\delta < x \neq 0 < \delta)$ *less (greater) than* $f(0, 0)$.

(2) *A positive number* δ' *must exist which is such that, if* y *be any number different from zero, and numerically less than* δ', *the upper (lower) limit of* $f(x, y)$, *for such constant value of* y, *and for all values of* x *for which* $-y \leqq x \leqq y$, *being* $f(\psi(y), y)$, *this upper (lower) limit is for every value of* $y\ (-\delta' < y \neq 0 < \delta)$ *less (greater) than* $f(0, 0)$.

It will be observed that, since $f(x, y)$ is assumed to be continuous with respect to x, and also with respect to y, the limit $f(x, \phi(x))$ is actually attained for some value $\phi(x)$ of y in the interval $(-x, x)$, and the limit $f(\psi(y), y)$ is actually attained for some value $\psi(y)$ of x in the interval $(-y, y)$. It is clear that, unless both the conditions stated in the theorem be satisfied, $f(0, 0)$ cannot be a proper maximum (minimum) of the function. If, for example, no such number as δ in (1) can be determined, there are points in every neighbourhood of $(0, 0)$ at which $f(x, y)$ is $\geqq (\leqq) f(0, 0)$.

The conditions are sufficient. For, if δ, δ' exist, the value of $f(x, y)$ at every point, except $(0, 0)$ within the neighbourhood the corners of which are the four points $(\pm\delta'', \pm\delta'')$ is less (greater) than $f(0, 0)$, where δ'' is the lesser of the two numbers δ, δ'.

The necessary and sufficient conditions that the function $f(x, y)$ may have an improper maximum (minimum) at $(0, 0)$ are similar to the above. In this case $f(x, \phi(x))$ must be less than, or equal to (greater than, or equal to) $f(0, 0)$ for all the values of x in the interval, and $f(\psi(y), y)$ must be less than, or equal to (greater than, or equal to) $f(0, 0)$ for all values of y in the interval. Further, corresponding to every positive number $\bar{\delta} < \delta$, there must be a value of $x\ (< \bar{\delta})$, for which $f(x, \phi(x)) = f(0, 0)$; or else a similar condition must hold for $f(\psi(y), y)$; or in both cases, the condition may be satisfied.

Other methods of determining whether $(0, 0)$ be a point at which there is a maximum or minimum of $f(x, y)$ will be dealt with in Chap. VI.

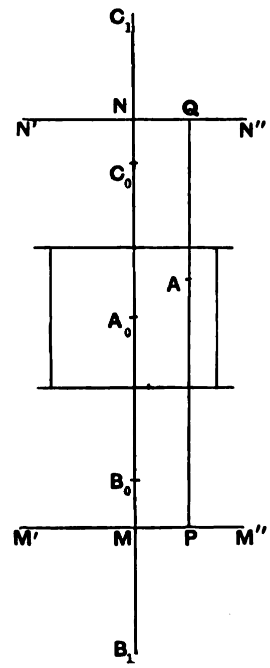
PROPERTIES OF A FUNCTION CONTINUOUS WITH RESPECT TO EACH VARIABLE.

243. Let a function $f(x, y)$, defined for all values of x and y in a continuous domain, be everywhere continuous with respect to y, and be also continuous with respect to x along each straight line parallel to the x-axis, and belonging to a set cutting the y-axis in an everywhere-dense set of points.

Let A be the point (x, y), and let BC be drawn with A as its middle point, parallel to the y-axis, and of length 2ρ. If $\omega(\rho)$ be the fluctuation of $f(x, y)$ in the interval BC, then $\omega(\rho)$ is a continuous function of ρ; and $\lim_{\rho=0} \omega(\rho) = 0$, since $f(x, y)$ is everywhere continuous with respect to y. Let σ be a fixed positive number, and let $\beta_\sigma(x, y)$ denote the upper limit of those values of ρ for which $\omega(\rho) \leqq \sigma$: thus $\omega(\rho) \leqq \sigma$, if $\rho \leqq \beta_\sigma(x, y)$; and $\omega(\rho) > \sigma$, if $\rho > \beta_\sigma(x, y)$.

The function $\beta_\sigma(x, y)$, thus defined for every point (x, y), is everywhere positive; and it will be shewn to be an upper semi-continuous function with respect to the two-dimensional continuum (x, y), in accordance with the definition in § 183.

Take $B_0A_0 = A_0C_0 = \beta_\sigma(x_0, y_0)$; and also $B_1B_0 = C_0C_1 = \frac{1}{2}\epsilon$, where ϵ is a fixed positive number. The fluctuation of $f(x, y)$ in B_1C_1 is greater than σ; let it be $\sigma + k$. If k_1 be a fixed positive number $< k$, two points M, N can be found in B_1C_1, such that



$$|f(M) - f(N)| > \sigma + k_1.$$

Moreover, these points M, N can be so chosen as to lie on two straight lines parallel to the x-axis, which belong to the set along each of which $f(x, y)$ is continuous with respect to x; this follows from the fact that this set of straight lines cuts B_1C_1 in an everywhere-dense set of points. Since $f(x, y)$ is continuous with respect to x, at each of the points M, N, two segments $M'M''$, $N'N''$, with M and N as their middle points, can be determined, so as to have equal lengths 2δ, and to be such that

$$|f(P) - f(M)| < \tfrac{1}{2}k_1,$$
$$|f(Q) - f(N)| < \tfrac{1}{2}k_1,$$

provided P be any point in $M'M''$, and Q be any point in $N'N''$.

From these inequalities and the former one, we deduce that

$$|f(P) - f(Q)| > \sigma.$$

Take the square of which A_0 is the centre, and of which the sides are parallel to the axes, and are at a distance from A_0 less than the smaller of the two numbers $\frac{1}{2}\epsilon$ and δ. If A be any point in this square, the distance of A from each of the straight lines $M'M''$, $N'N''$ is less than $\beta_\sigma(x_0, y_0) + \epsilon$. Through A let a straight line be drawn parallel to the y-axis, and mark off on it the segment of which A is the centre and of which the half-length is

$\beta_\sigma(x_0, y_0)+\epsilon$; this segment will cut $M'M''$ and $N'N''$, and therefore contains two points P, Q which are such that

$$|f(P)-f(Q)|>\sigma.$$

Therefore the fluctuation of $f(x, y)$ in this segment is $>\sigma$, and hence at the point A (x, y) we have $\beta_\sigma(x, y)<\beta_\sigma(x_0, y_0)+\epsilon$. A square having been determined with its centre at A_0, such that for every point in this square

$$\beta_\sigma(x, y)<\beta_\sigma(x_0, y_0)+\epsilon,$$

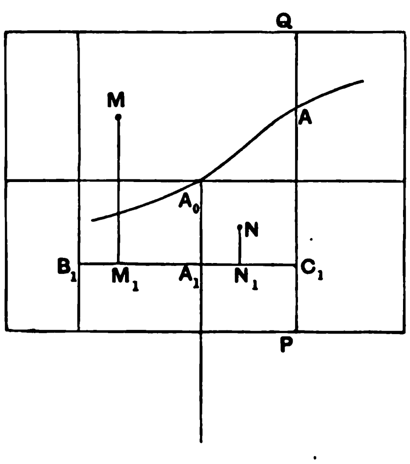
it follows that $\beta_\sigma(x, y)$ is an upper semi-continuous function at A_0 with respect to the two-dimensional domain (x, y).

244. Let us now consider the linear set C of points (x, y) defined by $y=\phi(x)$, where $\phi(x)$ is a continuous function of x. At each point of C, the function $\beta_\sigma(x, y)$ is defined, and has at every point a minimum relatively to the set C, the term being used in accordance with the definition given in § 180.

If, at a point A_0 (x_0, y_0) of C, the function $\beta_\sigma(x, y)$ have its minimum with respect to C positive, then we shall prove that the saltus of $f(x, y)$ at A_0, with respect to the two-dimensional domain (x, y), is $\leqq 2\sigma$. Let γ denote this minimum, and let γ_1 be a positive number $<\gamma$.

Let an interval $(x_0-\delta, x_0+\delta)$ on the line $y=y_0$, be so determined that in this interval

$$|\phi(x)-\phi(x_0)|<\tfrac{1}{2}\gamma_1.$$



This interval may, if necessary, be so reduced, that for all values of x in it,

$$\beta_\sigma\{x, \phi(x)\}>\gamma_1.$$

Describe the rectangle R, with A_0 as centre, the sides parallel to the axes of x and y being 2δ and γ_1 respectively. On every segment PQ of R, parallel to Oy, the fluctuation of the function is $\leqq\sigma$; for there is on PQ a point A of the set C, and the segment with centre A, and length $2\beta_\sigma(A)>2\gamma_1$, contains the whole segment PQ.

Taking a fixed positive number ϵ, an area surrounding A_0 can be determined, in which the fluctuation of $f(x, y)$ is $\leqq 2\sigma+\epsilon$. To effect this, take a point A_1 on the ordinate through A_0, and in the rectangle R, such that A_1 is a point of continuity of $f(x, y)$ with respect to x; then on the straight line $y=y_1$, take a segment B_1C_1 with centre A_1, and of length $2\delta'\leqq 2\delta$, such that the fluctuation of f in B_1C_1 is $<\epsilon$. Consider the rectangle R' contained in

R, such that the sides of R' are of lengths $2\delta'$ and γ_1 parallel to the axes, its centre being at A_0. The fluctuation of $f(x, y)$ in this rectangle is $< 2\sigma + \epsilon$. For if M, N be any two points in it, let M_1, N_1 be their projections on B_1C_1; then

$$|f(M) - f(M_1)| \leqq \sigma, \quad |f(N) - f(N_1)| \leqq \sigma, \quad |f(M_1) - f(N_1)| < \epsilon;$$

and from these inequalities we deduce that

$$|f(M) - f(N)| < 2\sigma + \epsilon.$$

Since this holds for every ϵ, the saltus of $f(x, y)$ at A_0 is $\leqq 2\sigma$. If, at a point A_0, the saltus of $f(x, y)$ be $> 2\sigma$, then at A_0 the minimum of β_σ with respect to C must be zero.

Since β_σ is positive at every point of C, and is an upper semi-continuous function of (x, y), it follows from the theorem of § 184, that in every arc D of the curve C, there exists an arc D_1 in which the minimum of β_σ is positive. Let us take a sequence $\sigma_1, \sigma_2, \ldots \sigma_n \ldots$ of positive decreasing numbers of which the limit is zero. It is then clear that in every arc D there exists a point where β_{σ_n} has its minimum with respect to C positive, for every σ_n. At this point the fluctuation of $f(x, y)$ with respect to the two-dimensional continuum (x, y) is $\leqq 2\sigma_n$, for all values of n, and is therefore zero. This point must be a point of continuity of $f(x, y)$ with respect to (x, y).

The following general theorem* has now been established:—

If $f(x, y)$ be a function of the two variables x, y which is everywhere continuous with respect to y, and is continuous with respect to x along straight lines parallel to the x-axis, which cut the y-axis in an everywhere-dense set of points, then in every portion of a curve $y = \phi(x)$, where $\phi(x)$ is a continuous function, there exist points at which $f(x, y)$ is continuous with respect to the two-dimensional domain (x, y).

It follows from this theorem that points of continuity exist in every area, that is $f(x, y)$ *is at most a point-wise discontinuous function.*

The whole of the reasoning above is applicable, if only those points of (x, y) are taken account of, which belong to a perfect set G. It thus appears that, under the conditions stated in the above theorem, $f(x, y)$ is a point-wise discontinuous function relatively to every perfect set G of points in (x, y). The points of continuity of $f(x, y)$ on the curve $y = \phi(x)$, are everywhere-dense with respect to every perfect set of points on the curve.

EXAMPLES.

1. If† $f(x, y)$ be a function which is everywhere continuous with respect to each of the variables x, y, then the points at which the saltus of $f(x, y)$ with respect to the two-dimensional continuum (x, y) is $\geqq \sigma$ form a set of points such that the projection of the set on either axis, by lines parallel to the other axis, is a non-dense set.

* Baire, *Annali di Mat.* Ser. III^a^, vol. III., 1899, p. 27. † Baire, *loc. cit.*, p. 94.

2. If* a function $f(x, y, z)$ of three variables x, y, z be everywhere continuous with respect to each variable, then $f(x, y, z)$ is at most a point-wise discontinuous function relatively to the three-dimensional continuum (x, y, z). Further, on every surface $x=\phi(y, z)$, where ϕ is continuous with respect to (y, z), the function $f(x, y, z)$ is at most a point-wise discontinuous function with respect to (y, z). The set of points at which the saltus of $f(x, y, z) \geqq \sigma$ may contain all the points of a continuous curve.

3. Let* $\phi(x, y)$ be a function which is continuous with respect to each of the variables x and y, and let $(0, 0)$ be a point of discontinuity of $\phi(x, y)$ with respect to (x, y). Define $f(x, y, z)$ by the condition $f(x, y, z)=\phi(x, y)$; then the function $f(x, y, z)$ is continuous with respect to each of the three variables, but every point on the z-axis is a point of discontinuity with respect to (x, y, z).

4. Let* $f(x, y, z)$ be a function which is constant along any straight line parallel to the straight line $x=y=z$, and is such that $f(x, y, 0)=\dfrac{xy(x-y)}{(x^2+y^2)^{\frac{3}{2}}}$, $f(0, 0, 0)=0$. This function is discontinuous at every point on the straight line $x=y=z$.

245. The methods developed by Baire of dealing with functions of two or more variables, in relation to the distribution of the points of discontinuity, have been applied by him to the consideration of the following three problems:—

(1) What must be the nature of a function $\phi(x)$, defined for $\alpha \leqq x \leqq \beta$, in order that a function $f(x, y)$ can exist which is defined for all points in the square $\alpha \leqq x \leqq \beta$, $\alpha \leqq y \leqq \beta$, and is continuous at every point with respect to x and with respect to y, and moreover is equal to $\phi(x)$ on the straight line $x=y$?

(2) What must be the nature of a function $\phi(x)$ defined for $\alpha \leqq x \leqq \beta$, in order that a function $f(x, y)$ can be defined for all points in the square $\alpha \leqq x \leqq \beta$, $\alpha \leqq y \leqq \beta$, and which shall satisfy the conditions that it is continuous with respect to (x, y) at every point for which $y > 0$, is continuous with respect to y at the points of $y=0$, and is equal to $\phi(x)$ when $y=0$?

(3) A function $f(x, y)$ is defined in the rectangle $\alpha \leqq x \leqq \beta$, $\gamma \leqq y \leqq \delta$, and is everywhere continuous with respect to y. Further, there is a set of parallels to the x-axis, along each of which $f(x, y)$ is continuous with respect to x; these parallels intersecting the straight line $x=\alpha$ in a set of points which is everywhere-dense in the interval (γ, δ). What is the nature of the function $f(x, y)$ on a continuous curve drawn in the rectangle?

The problems (1), (2) are particular cases of (3). It has been shewn above that a necessary condition satisfied by $f(x, y)$ in (3) is that it should be a point-wise discontinuous function relatively to every perfect set of points. That this condition is also sufficient, has been demonstrated by Baire in his memoir quoted above. A proof of this will be given for the case of problem (2), in Chapter VI, in connection with the theory of functions representable as the limits of sequences of functions.

* Baire, *loc. cit.*, p. 99.

THE REPRESENTATION OF A SQUARE ON A LINEAR INTERVAL.

246. Let a point of a square whose side is unity be denoted by (x, y), where $0 \leqq x \leqq 1$, $0 \leqq y \leqq 1$; and let t denote a point of a linear interval $(0, 1)$. An account has been given in § 58 of Cantor's method of establishing a $(1, 1)$ correspondence between the points of the square and those of the linear interval. Such a correspondence denotes functional relations $x = f(t)$, $y = \phi(t)$ between x, y as dependent variables, and t as an independent variable. It will be shewn however that no $(1, 1)$ relationship between the two sets of points can be a continuous representation*; *i.e.* it is impossible that the functions $f(t)$, $\phi(t)$ can be both continuous.

Let us assume that such a continuous representation can be defined. To any closed set of points $\{t\}$ in $(0, 1)$, there will correspond a closed set in the plane area. For if $t_1, t_2, \ldots t_n, \ldots$ be a convergent sequence of points t, of which t_ω is the limiting point, then the point $f(t_\omega)$, $\phi(t_\omega)$ is the limiting point of the set of points $(x_1, y_1)(x_2, y_2) \ldots (x_n, y_n) \ldots$ which correspond to $t_1, t_2, \ldots t_n, \ldots$ respectively; therefore to a closed set $\{t\}$ there corresponds a closed set $\{(x, y)\}$. Again, to a convergent sequence (x_1, y_1), $(x_2, y_2) \ldots$ of points in the plane area, there corresponds a set of points $t_1, t_2, \ldots$ in the linear interval, the latter of which has a limiting point t_ω which must correspond to (x_ω, y_ω); and since only one value of t corresponds to one set of values of (x, y), there can be only one such limiting point t_ω. Thus, to a closed set in the plane, there corresponds a closed set in the linear interval. Take two points t_1, t_2 in the interval $(0, 1)$; these points correspond to two points P_1, P_2 in the square area. To the closed linear interval (t_1, t_2) there corresponds a closed set S which contains the points P_1, P_2. It can be shewn that there are points other than P_1, P_2 on the frontier of S. Denote by $C(S)$ the set of those points of the square area which do not belong to S. Two points Q, R in the square can be determined, such that Q lies on the straight line P_1P_2, and R does not lie on this straight line; such that neither Q nor R coincides with P_1 or P_2, and such that one of the two belongs to S and the other to $C(S)$. The closed set consisting of the straight line QR contains points both of S and of $C(S)$; those points of S which lie on it form a closed set, and there must be one such point of S at least which is on the frontier of S; such a point may or may not coincide with Q or R. Since then S contains points on its frontier besides P_1 and P_2, we can take a point $\bar{t}$ within the linear interval (t_1, t_2) such that the point T in the square which corresponds to it is on the frontier of S. Since T is the limiting point of a sequence of points of $C(S)$, it follows that $\bar{t}$ must be the limiting point

* See Netto, "Beitrag zur Mannigfaltigkeitslehre," *Crelle's Jl.*, vol. LXXXVI.; also Loria, *Giorn. di Mat.*, vol. XXV., p. 97. In the proof given by these writers it is assumed that a closed curve corresponds to a linear sub-interval of (0, 1); this is not necessarily the case, for a non-dense closed set may correspond to the closed curve.

of a sequence of points all of which are external to the interval (t_1, t_2); and this is impossible. It has thus been established that:—

No continuous (1, 1) *correspondence can exist between all the points in a square and all the points in a linear interval.*

In particular, the correspondence shewn by Cantor to exist, must be discontinuous.

247. The reasoning of § 246 would be inapplicable if the correspondence $x = f(t)$, $y = \phi(t)$ were such that, to a given point (x, y) more than one point t may correspond, the functions $f(t)$, $\phi(t)$ being still one-valued continuous functions, so that if t be assigned, (x, y) is uniquely determined. In this case, the limiting point of the set of points external to the interval (t_1, t_2) would be not $\bar{t}$, but another value of t which also corresponds to the point T.

Peano* gave the first continuous correspondence of the kind just indicated, thus defining a continuous curve which passes through every point of the square at least once.

Let the points in the interval (0, 1) be expressed in the form

$$t = \cdot a_1 a_2 a_3 \ldots a_n \ldots,$$

in radix fractions in the ternary scale, so that each a is either 0, 1, or 2. Let $k(a)$ denote the number $2 - a$, so that $k(2) = 0$, $k(1) = 1$, $k(0) = 2$; and let $k^n(a)$ denote the result of performing this operation n times, so that $k^n(a)$ is a or $2 - a$, according as n is even or odd.

Let x, y be defined for a prescribed t by

$$x = \cdot b_1 b_2 b_3 \ldots, \quad y = \cdot c_1 c_2 c_3 \ldots,$$

the ternary scale being again employed; the numbers b, c being defined by the relations

$$b_1 = a_1, \quad b_2 = k^{a_2}(a_3), \ldots b_n = k^{a_2 + a_4 + \ldots + a_{2n-2}}(a_{2n-1}),$$

$$c_1 = k^{a_1}(a_2), \quad c_2 = k^{a_1 + a_3}(a_4), \ \ldots \ c_n = k^{a_1 + a_3 + \ldots + a_{2n-1}}(a_{2n});$$

thus b_n is equal to a_{2n-1} or to $2 - a_{2n-1}$, according as $a_2 + a_4 + \ldots + a_{2n-2}$ is even or odd.

The numbers t may be divided into two classes:—

(1) Those, other than 0 or 1, which are capable of a double representation

$$t = \cdot a_1 a_2 a_3 \ldots a_n 2\,2\,2 \ldots = \cdot a_1 a_2 \ldots \overline{a_n + 1}\, 0\,0\,0 \ldots.$$

(2) Those which have a single representation only.

If t be a number of the second class, x and y are uniquely defined. If t be a number of the first class

$$t = \cdot a_1 a_2 \ldots a_n 2\,2\,2 \ldots \equiv \cdot a_1 a_2 \ldots \overline{a_n + 1}\, 0\,0\,0 \ldots,$$

* "Sur une courbe, qui remplit toute une aire plane," *Math. Ann.*, vol. XXXVI., 1890.

let $\cdot b_1 b_2 b_3 \ldots$, $\cdot b_1' b_2' b_3' \ldots$ denote the numbers obtained by applying the definition of x to the two modes of representation of t. If n is even, say $2m$, it is clear that

$$b_1 = b_1', \ b_2 = b_2', \ldots b_m = b_m';$$

also
$$b_{m+1} = k^{a_2 + a_4 + \ldots + a_{2m}} 2, \qquad b'_{m+1} = k^{a_2 + a_4 + \ldots + a_{2m} + 1} 0,$$
$$b_{m+2} = k^{a_2 + a_4 + \ldots + a_{2m} + 2} 2, \qquad b'_{m+2} = k^{a_2 + a_4 + \ldots + a_{2m} + 1} 0;$$
$$\ldots\ldots\ldots\ldots \qquad \ldots\ldots\ldots\ldots$$

hence
$$b_{m+1} = b'_{m+1}, \ b_{m+2} = b'_{m+2}, \ldots;$$

and thus x has the same value whichever of the two forms for t is employed; the case in which n is odd may be similarly treated.

The same result can readily be shewn to hold for y. Therefore, corresponding to any assigned t, x and y are uniquely determined.

Next, let us suppose x and y to be assigned. We have

$$a_1 = b_1, \ a_2 = k^{b_1}(c_1), \ a_3 = k^{c_1}(b_2), \ a_4 = k^{b_1 + b_2}(c_2), \ \ldots$$
$$a_{2n-1} = k^{c_1 + c_2 + \ldots + c_{n-1}}(b_n), \qquad a_{2n} = k^{b_1 + b_2 + \ldots + b_n}(c_n);$$

for, if $p = k^r(q)$, then $p + q$ is an even number.

In case x, y are both of the second class, t is uniquely determined.

If x is of the first class, and y of the second; let

$$x = \cdot b_1 b_2 \ldots b_n 2 2 2 \ldots = \cdot b_1 b_2 \ldots \overline{b_n + 1}\, 0\, 0 \ldots,$$
$$y = \cdot c_1 c_2 \ldots c_n c_{n+1} \ldots,$$

and let the two values of t be denoted by $\cdot a_1 a_2 a_3 \ldots$, $\cdot a_1' a_2' \ldots$.

It is clear that
$$a_1 = a_1', \ a_2 = a_2' \ldots a_{2n-1} = a'_{2n-1};$$

also $a_{2n} = k(a'_{2n})$, $a_{2n+1} = k^{c_1 + c_2 + \ldots + c_n}(b_n)$, $a'_{2n+1} = k^{c_1 + c_2 + \ldots + c_n}(b_n + 1)$; thus a_{2n+1}, a'_{2n+1} are not identical, although a_{2n}, a_{2n}' will be so if each is unity. It is thus seen that t has two distinct values corresponding to one point (x, y) when x is a number of the first class, and y is of the second class. It can be shewn in a similar manner that there are four points t corresponding to a single point (x, y) such that x, y are both numbers of the first class.

The correspondence is continuous. For if t, t' are identical as regards the first $2n$ figures, x and x' are identical as regards their first n figures, and the same is true of y and y'.

The curve which has thus been defined is a continuous curve which passes through each point in the square at least once; there is an everywhere-dense enumerable set of points through each of which the curve passes twice, and another everywhere-dense enumerable set of points through each of which it passes four times; through each of the remaining unenumerable set of points, the curve passes once only.

The plane measure of an arc of Peano's curve which corresponds to an interval (t_0, t_1) is not zero, *i.e.* the area which a number of rectangles enclosing all the points of the arc have in common has a lower limit greater than zero.

The two continuous functions $f(t)$, $\phi(t)$, which define x, y as functions of t, possess for no value of t definite differential coefficients, and are perhaps the simplest examples of continuous non-differentiable functions.

248. It might at first sight appear that a curve having the same properties as that of Peano might have been defined by restricting $t = \cdot a_1 a_2 \ldots$ to be such that an infinite number of digits other than 0 are present, and then defining x, y by

$$x = \cdot a_1 a_3 a_5 \ldots, \quad y = \cdot a_2 a_4 a_6 \ldots.$$

If however the double representation of x, y were not restricted, as in the case of t, there would be no value of t corresponding to, say,

$$x = \cdot 1000\ldots, \quad y = \cdot 2000\ldots.$$

If (x, y) were on the other hand so restricted, there would be no values of (x, y) corresponding, for example, to

$$t = \cdot 111010101\ldots.$$

It thus appears that some such rule as that given by Peano is necessary to obviate the difficulty caused by the double representation of a certain class of rational numbers, in a given scale.

The method may easily be extended to obtain a continuous correspondence between the points in a cube and those in a linear interval.

A somewhat different method of establishing correspondence between the points of the square, and those of the linear interval, is the following* :—

Let t_1 denote one of the perfect set of points defined by

$$t_1 = \frac{a_1}{3} + \frac{a_2}{3^2} + \frac{a_3}{3^3} + \ldots,$$

when every a is either 0 or 2. For such a point t_1, x and y may be defined by

$$x = \frac{1}{2}\left(\frac{a_1}{2} + \frac{a_3}{2^2} + \frac{a_5}{2^3} + \ldots\right),$$

$$y = \frac{1}{2}\left(\frac{a_2}{2} + \frac{a_4}{2^2} + \frac{a_6}{2^3} + \ldots\right).$$

A point t which does not belong to the perfect set is interior to one of the

* See Lebesgue, *Leçons sur l'intégration*, p. 44.

complementary intervals (t_1', t_1'') of the set; in such an interval we may define x, y as linear functions of t, thus

$$x = x' + \frac{x' - x''}{t_1' - t_1''}(t - t_1'),$$

$$y = y' + \frac{y' - y''}{t_1' - t_1''}(t - t_1'),$$

where (x', y'), (x'', y'') correspond to t_1', t_1'' respectively.

249. A method of constructing a continuous curve which fills a square has been given in a geometrical form by Hilbert*.

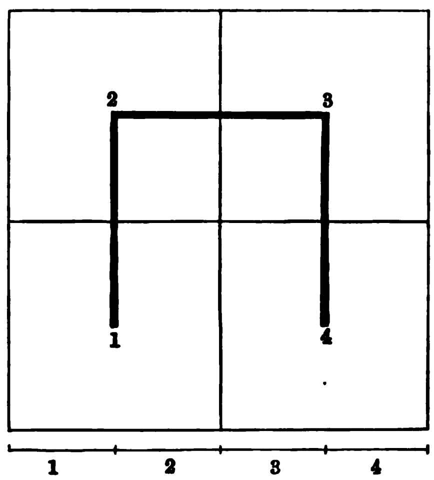


Fig. 1.

Divide the interval (0, 1) into four equal parts, and number them in order as 1, 2, 3, 4. Then divide the square into four equal parts, as in Fig. 1, and number them 1, 2, 3, 4, to correspond with the segments of the linear interval. Next divide each segment of the straight line into four equal parts, and each of the four squares into four equal parts as in Fig. 2. The sixteen squares so formed are then numbered in order so that each square has one side in common with the one next in order; the squares then correspond with the segments numbered in the same way. At the next stage there are (Fig. 3) 64 squares corresponding to 64 segments of the interval (0, 1). Proceeding in this manner indefinitely, any point of (0, 1) is determined by the intervals of the successive set of sub-divisions in which it lies. The corresponding point in the square area is determined by the succession of squares each containing the next in which it lies. The curve is thus determined as the limit of a sequence of polygons denoted by the thickened lines in the figures. The curve thus obtained is continuous, but has no tangent.

* See *Math. Annalen*, vol. xxxviii., p. 459.

Hilbert remarks that if the interval (0, 1) be taken as a time interval, a kinematical interpretation of the functional relation between the curve and the segment is that a point may move so that in a finite time it passes through every point of the square area.

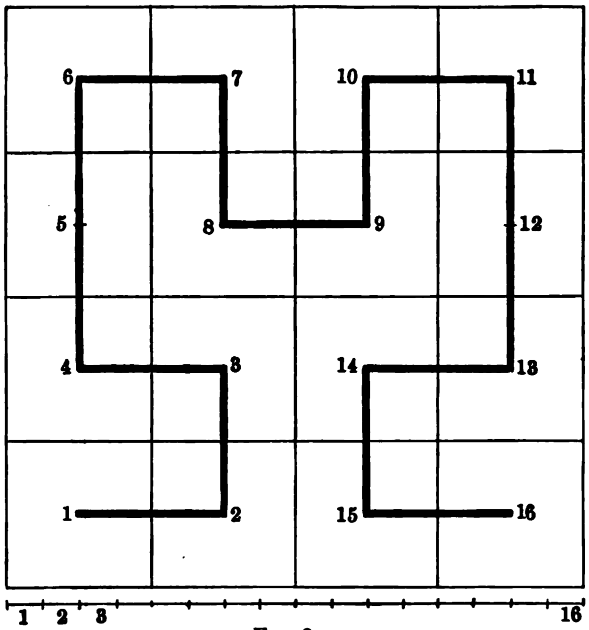


Fig. 2.

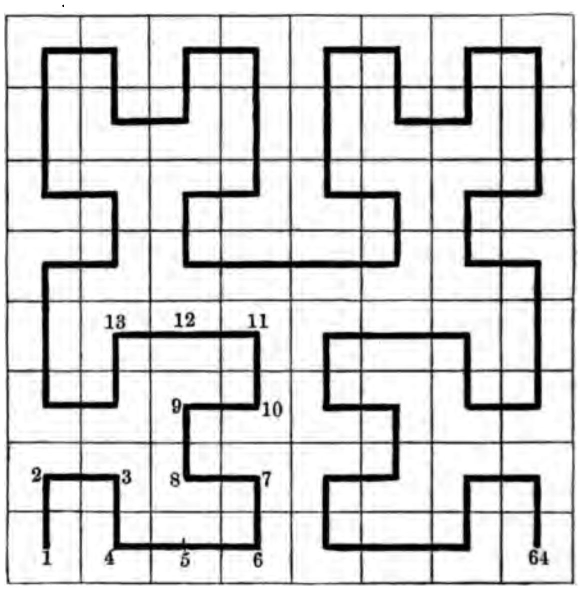
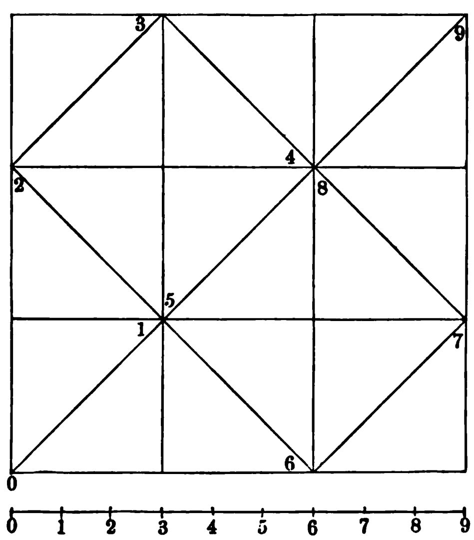


Fig. 3.

Continuous curves of this kind can be constructed by any method by which an everywhere-dense enumerable set of points in the square can be made to correspond with a similar set of points in the linear interval; provided the functional relation $x = f(t)$, $y = \phi(t)$, in such correspondence, is uniformly continuous. For, when this condition is satisfied, the functions obtained by the method of extension of $f(t)$, $\phi(t)$ to the remaining points of (0, 1) as secondary points (see § 225) will yield a correspondence of all the points of the square with those of the linear interval, of the required character.

Another method differing from that of Hilbert has been given by Moore* and by Schönflies†.

Let m be an uneven number (in the figure, $m = 3$); divide the linear interval (0, 1) into m^2 equal parts, and also the square into m^2 equal parts. Let these



squares be passed through by a polygonal line, of which the sides are diagonals of the squares, as in the figure; in this manner the squares are arranged in order 1, 2, 3, ... m^2, and are placed into correspondence with the segments bearing the same numbers. At the same time the end-points of a diagonal so traversed are made to correspond with the end-points of a segment of the linear interval. Thus $m^2 + 1$ points in the linear interval are placed in correspondence with points in the square, so that to each of the $m^2 + 1$ points of the linear interval there is one point in the square; but the converse is not the case. Next, divide each of the m^2 linear intervals into m^2

* *Trans. Amer. Math. Soc.*, vol. I., p. 77.
† *Bericht über die Mengenlehre*, p. 121.

equal parts, and the corresponding squares into m^2 equal parts; then construct as before a polygon traversing diagonals of all the m^4 squares, making their end-points correspond to the end-points of the corresponding m^4 parts of the linear interval. Proceeding in this manner, we gradually place points in the square, consisting of an everywhere-dense enumerable set, into correspondence with a set in the linear interval which possesses the same property; and the functional relation so set up is uniformly continuous. The definition of the functions for the whole linear interval is then obtained, as explained above, by the method of extension. The case $m = 3$, corresponds to Peano's analytical method. In the method of Moore and Schönflies, the curve is determined as the limit of a sequence of polygons inscribed in the curve. In Hilbert's method the polygons which approximate to the form of the curve are not inscribed in the curve, but are otherwise determined.

CHAPTER V.

INTEGRATION.

250. The fundamental operation of the calculus, known as integration, regarded from one point of view consists essentially in the determination of the limit of the sum of a finite series of numbers, as the number of terms of the series is indefinitely increased, whilst the numerically greatest of the individual terms of the series approaches the limit zero. The laws which regulate the specification of the terms of the series must be supposed, in any given instance, to be assigned, and to be of such a character that the limit in question exists. It is in this form that the problem of integration naturally presents itself in ordinary problems of a geometrical character, such as the determination of lengths, areas, volumes, &c. The method of integration, so regarded, has its origin in the method of exhaustions employed by the Greek geometers, and was developed later in forms of which the exactitude depended at various epochs upon the stage which the development of Analysis in general had reached. In the hands of Cauchy, Dirichlet, and Riemann the definition of the definite integral attained to the exact arithmetic form in which it is employed in modern analysis; and in fact the definition given by Riemann, which is now held to be fundamental in the calculus, leaves nothing to be desired as regards precision. Riemann gave not only a precise definition, but also a necessary and sufficient condition, for the existence of the definite integral. Although a more general definition of integration has recently been developed by Lebesgue, in accordance with which classes of functions are integrable, which are not so in accordance with the definition of Riemann, the latter is the definition which lies at the base of almost all the developments of the theory of integration that have been made during the last half century, and will therefore be adopted for full treatment in the present Chapter. An account will however be given of the recent more general theory due to Lebesgue.

Integration has also usually been regarded as the operation inverse to that of differentiation; and the fundamental theorem of the Integral Calculus formulates the relation of this mode of regarding integration with the one referred to above. Many important investigations are concerned with the relation between these two modes of regarding integration, with the establishment of the fundamental theorem, and with an examination of the limitations to which it is subject.

THE DEFINITE INTEGRALS OF LIMITED FUNCTIONS.

251. Let $f(x)$ be a limited function, defined for the continuous domain (a, b), where $b \gtrless a$; so that there exists an upper limit U and a lower limit L of the functional values in the whole interval. Let the interval (a, b) be divided into any n_1 sub-intervals $\delta_1^{(1)}, \delta_2^{(1)}, \delta_3^{(1)}, \dots \delta_{n_1}^{(1)}$, so that

$$\delta_1^{(1)} + \delta_2^{(1)} + \dots + \delta_{n_1}^{(1)} = b - a,$$

and let Δ_1 be the greatest of these sub-intervals. Let these sub-intervals be further sub-divided in any manner so that the whole interval (a, b) then consists of n_2 sub-intervals $\delta_1^{(2)}, \delta_2^{(2)} \dots \delta_{n_2}^{(2)}$ whose sum is $b - a$, and the greatest of which is Δ_2; let further sub-divisions of these sub-intervals be made, and so on continually, so that at any stage of the process the interval (a, b) is divided into n_m sub-intervals, $\delta_1^{(m)}, \delta_2^{(m)}, \dots \delta_{n_m}^{(m)}$, the greatest of which is Δ_m. If this system of continual sub-division be made in any manner whatever, which is such that the sequence $\Delta_1, \Delta_2, \dots \Delta_m, \dots$ has the limit zero, we shall, as in § 193, speak of it as a convergent system of sub-divisions of the interval (a, b). Let $M(\delta_s^{(m)})$ denote any number whatever which is so chosen as to be not greater than the upper limit of the function $f(x)$ in the closed interval $\delta_s^{(m)}$, and so as to be not less than the lower limit of $f(x)$ in the same interval; and consider the sums

$$S_1 = \delta_1^{(1)} M(\delta_1^{(1)}) + \delta_2^{(1)} M(\delta_2^{(1)}) + \dots + \delta_{n_1}^{(1)} M(\delta_{n_1}^{(1)}),$$
$$S_2 = \delta_1^{(2)} M(\delta_1^{(2)}) + \delta_2^{(2)} M(\delta_2^{(2)}) + \dots + \delta_{n_2}^{(2)} M(\delta_{n_2}^{(2)}),$$
$$\dots$$
$$\dots$$
$$S_m = \delta_1^{(m)} M(\delta_1^{(m)}) + \delta_2^{(m)} M(\delta_2^{(m)}) + \dots + \delta_{n_m}^{(m)} M(\delta_{n_m}^{(m)}).$$

If the sequence $S_1, S_2, \dots S_m, \dots$ be convergent and have the same number S for limit whatever convergent system of sub-divisions of (a, b) be employed, and however the numbers $M(\delta_s^{(m)})$ be chosen, subject only to their limitation in relation to the upper and lower limits of $f(x)$ in the intervals $\delta_s^{(m)}$, then the function $f(x)$ is said to be integrable in the interval (a, b), and the number S defines the value of its integral. This integral, when the limit S exists, is denoted by $\int_a^b f(x)\,dx$.

It will be observed that $M(\delta)$ is not necessarily the value of $f(x)$ at any point in the interval δ; for all that is necessary is that it should not be greater than the upper limit, nor less than the lower limit, of $f(x)$ in the interval δ. In this respect the definition is a slight generalization of that given by Riemann*, who restricted $M(\delta)$ to have the value of $f(x)$ at some point in the interval δ.

The definition of a definite integral, of which Riemann's definition is a

* *Werke*, 2nd ed. p. 239.

development, was given by Cauchy*, for the case of a continuous function. Cauchy's definition is in fact that which arises when $M(\delta)$ is in every case restricted to be the functional value at one end of the interval δ; thus it may be expressed by

$$\int_a^b f(x)\,dx = \lim\,[(x_1-a)f(a)+(x_2-x_1)f(x_1)+\ldots+(b-x_n)f(x_n)],$$

where $a, x_1, x_2, \ldots x_n, b$ are the end-points of the sub-divisions, and the limit is determined under the same conditions as have been stated above.

252. The investigation of the necessary and sufficient conditions that the integral of $f(x)$ in (a, b), as above defined, may exist, is considerably simplified by the introduction of the notions of the upper† and lower integrals of the function $f(x)$ in the interval (a, b).

If, in the successive sums which are formed corresponding to a convergent system of sub-divisions of (a, b), we identify every number $M(\delta)$ with the upper limit $U(\delta)$ of the function in the interval δ, it can be shewn that for any limited function whatever, the sequence of numbers

$$\delta_1^{(1)}U(\delta_1^{(1)})+\delta_2^{(1)}U(\delta_2^{(1)})+\ldots+\delta_{n_1}^{(1)}U(\delta_{n_1}^{(1)})=\Sigma_1,$$
$$\delta_1^{(2)}U(\delta_1^{(2)})+\delta_2^{(2)}U(\delta_2^{(2)})+\ldots+\delta_{n_2}^{(2)}U(\delta_{n_2}^{(2)})=\Sigma_2,$$
$$\ldots$$
$$\delta_1^{(m)}U(\delta_1^{(m)})+\delta_2^{(m)}U(\delta_2^{(m)})+\ldots+\delta_{n_m}^{(m)}U(\delta_{n_m}^{(m)})=\Sigma_m$$

has a definite limit when m is indefinitely increased, which is independent of the particular convergent system of sub-intervals. This limit is called the *upper integral* of $f(x)$ in the interval (a, b), and may be denoted by

$$\overline{\int_a^b} f(x)\,dx.$$

A similar theorem holds if $M(\delta)$ be in every case identified with the lower limit of $f(x)$ in the interval, the corresponding sum converging to a number which is also independent of the particular convergent system of sub-intervals chosen. This limit is then termed the *lower integral* of $f(x)$ in (a, b), and is denoted by $\underline{\int_a^b} f(x)\,dx$.

To prove that the upper integral of a limited function always exists, we observe that when any sub-interval is subdivided the upper limit in no one of the sub-divisions can be greater than in the original sub-interval, and consequently Σ_{m+1} cannot be greater than Σ_m. It thus appears that $\Sigma_1, \Sigma_2, \ldots \Sigma_m, \ldots$

* *Journal de l'École Polytechnique*, cah. 19 (1823), pp. 571 and 590.

† The upper integral (oberes Integral) and the lower integral (unteres Integral) are named by Jordan, "l'intégrale par excès," and "l'intégrale par défaut" respectively; see *Cours d'Analyse*, vol. I, p. 34. They were introduced by Darboux, *Annales de l'école normale*, ser. 2, vol. IV, and also by Thomae, *Einleitung*, p. 12, and by Ascoli, *Atti di Lincei*, ser. 2, vol. II, 1875, p. 863.

form a sequence of numbers which do not increase, and moreover none of them is less than $L(b-a)$; consequently they form a convergent sequence of which we may denote the limit by N. It must now be shewn that N is independent of the particular convergent system of sub-divisions. Suppose, if possible, that another system of sub-divisions leads to another limit N'; we may without loss of generality suppose that $N' < N$. Take a system of intervals $\epsilon_1, \epsilon_2, \dots \epsilon_s$ belonging to the second system of sub-divisions, where we may suppose s to be so great that the sum for this system is $< N' + \zeta$, where ζ is an arbitrarily small number, and we choose it so that $N' + \zeta < N$. Let $n_m > s$, and suppose the two sets of sub-divisions

$$\delta_1^{(m)}, \delta_2^{(m)}, \dots \delta_{n_m}^{(m)}, \qquad \epsilon_1, \epsilon_2, \dots \epsilon_s$$

to be superimposed, so that (a, b) is divided up by all the points which are end-points of sub-intervals of either set; the new division of (a, b) may be regarded as a continuation of either set of sub-intervals into further sub-division. Since $s < n_m$, at most $s-1$ of the n_m intervals are divided by introducing the points belonging to the ϵ, and the diminution thus produced in Σ_m is less than or equal to $(s-1)\Delta_m(U-L)$, and thus the new sum for the combined sub-divisions is $\geqq \Sigma_m - (s-1)\Delta_m(U-L)$. Now m can be chosen so great that $\Delta_m < \dfrac{\eta}{(s-1)(U-L)}$, where η is an arbitrarily chosen positive number as small as we please; and if this be done the sum for the combined sub-divisions is $> \Sigma_m - \eta > N - \eta$. Again, since the same sum may be regarded as belonging to a further sub-division of the intervals $\epsilon_1, \epsilon_2, \dots \epsilon_s$, it is $< N' + \zeta$. It is now clear that, since η can be chosen so that $N - \eta > N' + \zeta$, the sum for the combined system of sub-divisions cannot be both $> N - \eta$ and less than $N' + \zeta$; and it is thus impossible that N and N' should be unequal: therefore the limiting sum which has been shewn to exist for any prescribed system of sub-divisions has the same value for all such systems. The existence of the lower integral may be proved by similar reasoning, or is immediately deducible from the existence theorem for the upper integral by considering the function $-f(x)$.

253. It has now been shewn that a limited function $f(x)$, defined for the continuous interval (a, b), always possesses an upper and a lower integral in the interval. The necessary and sufficient condition that $f(x)$ should possess an integral as defined in § 251 is that the upper and lower integrals in the interval be equal. That this condition is necessary follows at once from the fact that all the numbers $M(\delta)$ may be made identical with $U(\delta)$, or all may be made identical with the lower limits $L(\delta)$ of the functions in the intervals δ; and that the condition is sufficient follows from the fact that S_m lies between $\sum_{s=1}^{s=n_m} \delta_s^{(m)} U(\delta_s^{(m)})$ and $\sum_{s=1}^{s=n_m} \delta_s^{(m)} L(\delta_s^{(m)})$, and thus that when the two latter sums have the same limit, that limit is also the limit of S_m.

The necessary and sufficient condition that $f(x)$ may be integrable in the interval (a, b) may now be expressed as follows:—Let $D(\delta^{(m)})$ denote the fluctuation $U(\delta^{(m)}) - L(\delta^{(m)})$ of the function in the interval $\delta^{(m)}$; then it must be possible to define a convergent system of sub-divisions of the interval (a, b) such that, if at any stage these sub-divisions are denoted by $\delta_1^{(m)}, \delta_2^{(m)}, \ldots \delta_{n_m}^{(m)}$, the sum*

$$\delta_1^{(m)} D(\delta_1^{(m)}) + \delta_2^{(m)} D(\delta_2^{(m)}) + \ldots + \delta_{n_m}^{(m)} D(\delta_{n_m}^{(m)})$$

should have the limit zero, as m is increased indefinitely.

That this zero limit exists is equivalent to saying that, corresponding to any arbitrarily small positive number ϵ, a number m can be found such that the absolute value of $\sum_{s=1}^{s=n_m} (\delta_s^{(m)}) D(\delta_s^{(m)})$, or of $(b-a)M$, where M is a certain mean of the numbers $D(\delta_s^{(m)})$, for this value of m and for all greater values of m, shall be less than ϵ.

254. The necessary and sufficient condition for the existence of $\int_a^b f(x)\,dx$, may be stated in a somewhat more convenient form, as follows:—

If any convergent system of sub-divisions of the interval (a, b) be taken, then, corresponding to any arbitrarily chosen positive number k, the sum of those sub-intervals of (a, b) in which the fluctuation of $f(x)$ is greater than or equal to k, must, as the successive sub-division advances, become arbitrarily small, and must then have the limit zero.

To see that the condition so stated is sufficient, we observe that, if $s^{(m)}$ be the sum of those sub-intervals of $\delta_1^{(m)}, \delta_2^{(m)}, \ldots \delta_{n_m}^{(m)}$ in which the fluctuation is $\geqq k$, then

$$\sum_{t=1}^{t=n_m} \delta_t^{(m)} D(\delta_t^{(m)}) \leqq s^{(m)}(U-L) + k(b-a-s^{(m)}).$$

Since $s^{(m)}$ has the limit zero as m is increased indefinitely, the limit of $\Sigma\delta_t^{(m)} D(\delta_t^{(m)})$ is $\leqq k(b-a)$; and as k is arbitrary, the limit must be zero.

To shew that the condition is necessary, we observe that

$$\sum_{t=1}^{t=n_m} \delta_t^{(m)} D(\delta_t^{(m)}) \geqq ks^{(m)} + (b-a-s^{(m)})\overline{D} \geqq ks^{(m)},$$

where $\overline{D}$ is the least of the fluctuations in all the sub-intervals. Unless therefore $s^{(m)}$ has the limit zero it is impossible that $\Sigma\delta^{(m)} D(\delta^{(m)})$ can have the limit zero.

Another form of the condition for the existence of $\int_a^b f(x)\,dx$ which is for many purposes more convenient than the above forms of statement, involves the saltus or measure of discontinuity at points of the interval instead of the fluctuations in sub-intervals; it may be stated as follows:—

* Riemann's *Werke*, 2nd ed. p. 240.

The necessary and sufficient condition that the limited function $f(x)$ may be integrable in the interval (a, b) is that, for any value whatever of the positive number k, those points of the interval at which the saltus σ is $\geqq k$ form a set of points of zero content.

To see that the condition is necessary, let any system of sub-divisions of (a, b) be taken; then the sum of the products of the sub-intervals multiplied by the corresponding fluctuations is greater than k times the sum of those sub-intervals which contain points of the set for which $\sigma \geqq k$; unless therefore the sum of these sub-intervals have the limit zero as the sub-division advances, it is impossible that the sum of the products of sub-intervals and fluctuations should have the limit zero. To shew that the condition is sufficient, we observe that if the content of the set of points $\sigma \geqq k$ be zero, all these points can be included in a definite number of sub-intervals whose sum is less than the arbitrarily small number ϵ, so that all the points of the set are interior points of these sub-intervals; and the rest of the interval (a, b) consists of a definite number of sub-intervals whose sum is greater than $b - a - \epsilon$, and at every point of which $\sigma < k$. Consider one of these latter sub-intervals δ. In accordance with the theorem established in § 185, δ can be divided into a definite number of parts in each of which the fluctuation is less than k. Since the same reasoning applies to every sub-interval δ, therefore the whole interval (a, b) can be divided into a definite number of sub-intervals such that the sum of those in which the fluctuation is $\geqq k$ is less than ϵ, and this however small ϵ may be; and this is the condition of integrability established above.

The most succinct form in which the condition of integrability of a limited function may be stated is the following*:—

The necessary and sufficient condition that a limited function defined for a given interval may be integrable is that the points of discontinuity of the function form a set of measure zero.

For if $k_1, k_2, \dots k_n, \dots$ be a sequence of diminishing positive numbers which converges to the limit zero, and $G_1, G_2, \dots G_n, \dots$ be the closed sets of points at which the saltus of the function is $\geqq k_1, \geqq k_2, \dots \geqq k_n, \dots$, then the set of all the discontinuities of the function is the limit of the set G_n, when n is indefinitely increased, and this set must, in accordance with the theorem of § 88, have the measure zero, since G_n has the measure zero, for every value of n.

This condition is equivalent to the condition that every closed set, contained in the set of points of discontinuity of the function, may have content zero†.

* Lebesgue, *Annali di Mat.* ser. 3, vol. VII, p. 254.

† See W. H. Young, *Quarterly Journal of Math.* vol. XXXV, p. 190. See also Hobson, *Quarterly Journal*, vol. XXXV, p. 208.

PARTICULAR CASES OF INTEGRABLE FUNCTIONS.

255. The following classes of limited functions satisfy the condition of integrability which has been expressed in various forms above.

(1) All functions which are continuous in the intervals for which they are defined.

(2) All functions with only a finite number of discontinuities, or with any enumerable set of discontinuities

(3) Monotone functions, and all functions with limited total fluctuation.

For, as has been shewn in § 194, the points of discontinuity of a function with limited total fluctuation form an enumerable set.

(4) Generally, every point-wise discontinuous function which is such that the closed set of points for which the saltus is $\geqq k$ has content zero, whatever positive value k may have.

Dini* has given the theorem that *a function is integrable, if at all points where the discontinuity is of the second kind, it is so for all such points only on one and the same side of the point; and at these points the function may be continuous on the other side, or may have ordinary discontinuities on that side.* In particular, *any function which has only ordinary discontinuities is integrable.*

To prove this we observe that it has been proved in § 189, that, for such a function, the set of points for which the saltus is $\geqq k$ has content zero, whatever positive value k may have. Therefore the condition of integrability is satisfied.

Riemann's definition of an integral, and the condition for the existence of the integral, are applicable, without essential change, to the case of a function which, for particular values of the variable, has indeterminate functional values lying, in the case of each such point, between finite limits of indeterminacy. At each point of indeterminacy of the function, it is immaterial whether the function be capable of having all, or only some, values between the limits of indeterminacy; thus there is no loss of generality, if the function be regarded as having two values only at each such point, viz. the two limits of indeterminacy at the point. In estimating the fluctuation of the function in a prescribed interval, the upper limit is found by taking the upper limits of indeterminacy of the function at the special points as functional values at those points, whilst the lower limit is found by taking the lower limits of indeterminacy at the special points as the functional values at those points. As in the case of a function which is everywhere single-valued, the saltus at any point is defined as the limit of the fluctuation

* See *Grundlagen*, p. 335.

in a neighbourhood of the point, when that neighbourhood is diminished indefinitely. The conditions of integrability are exactly the same as for a function which is everywhere single-valued, viz. that the function be limited in its domain and that the set of points of discontinuity of the function must have zero measure.

PROPERTIES OF THE DEFINITE INTEGRAL.

256. We proceed to consider the properties of the integral $\int_a^b f(x)\,dx$ of a limited function $f(x)$, defined for the interval (a, b), and such that the condition for the existence of the integral is satisfied.

(1) *The integral* $\int_b^a f(x)\,dx$ *exists and has the value* $-\int_a^b f(x)\,dx$.

For the former integral when it exists is defined by means of the limit of $\Sigma\delta M(\delta)$, where δ is one of a set of finite intervals into which (b, a) is divided; any such interval differs from a corresponding interval in (a, b) only in sign, and the numbers $M(\delta)$ may be taken to be the same for corresponding intervals in the two cases. It is thus clear that the existence of the one limit follows from that of the other, and that they differ only in sign.

(2) *If* $f(x)$ *be integrable in* (a, b), *so also is* $|f(x)|$, *and*

$$\left|\int_a^b f(x)\,dx\right| \leqq \int_a^b |f(x)|\,dx.$$

For the fluctuation of $|f(x)|$ in any interval δ cannot exceed that of $f(x)$ in the same interval; hence, if $\Sigma D\delta$ for a convergent sequence of sub-intervals have the limit zero when D is the fluctuation of $f(x)$ in δ, it has also the limit zero when D denotes the fluctuation of $|f(x)|$; and thus the latter function is integrable. Again, U the upper limit of $f(x)$ in δ cannot numerically exceed U', the upper limit of $|f(x)|$ in the same interval; thus $|\Sigma U\delta| \leqq \Sigma U'\delta$, and hence the absolute value of the limit of $\Sigma U\delta$ is $\leqq$ that of $\Sigma U'\delta$.

(3) *If the values of the integrable function* $f(x)$ *be arbitrarily altered at each point of a measurable set of points* G, *the new function* $\phi(x)$ *so obtained is integrable, provided it be limited, and the measure of the derivative* G' *of the set be zero.*

For the only points of discontinuity of $\phi(x)$ which are not points of discontinuity of $f(x)$ are points of G or of G', and therefore form a set of measure zero; hence all the discontinuities of $\phi(x)$ form a set of points of zero measure, and $\phi(x)$ is therefore integrable provided it be limited. In particular, the theorem holds for any reducible set G.

Also, if $\phi(x) = f(x)$, *at all points belonging to a set which is everywhere-dense in* (a, b), *then, provided* $\phi(x)$ *be integrable, its integral is identical with that of* $f(x)$.

For, in the finite sum $\Sigma\delta M(\delta)$, we may take the value of $M(\delta)$ in any interval δ to be one of the values which the two functions $f(x)$, $\phi(x)$ have in common in that interval; hence the sums may all be chosen so as to be the same for the two functions. Thus, if the functions be both integrable, their integrals are identical.

(4) *A function $f(x)$ which is integrable in (a, b) is also integrable in any interval (α, β) contained in (a, b).*

For the measure of the set of points of discontinuity of $f(x)$ in (a, b) being zero, the measure of the set of those points of discontinuity which are in (α, β) is also zero, and thus the function is integrable in (α, β).

If c is any point in (a, b), we have

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

For the two integrals on the right-hand side both exist; also a convergent sequence of sets of sub-divisions of (a, b) can be so chosen that the point c is always an end-point of two of the sub-divisions. If this be done, the sum $\Sigma\delta M(\delta)$ for (a, b) may be divided into two parts, one of which contains all the intervals on the left of the point c, and the other all those on the right of that point; thus $\Sigma\delta M(\delta) = \Sigma_1\delta M(\delta) + \Sigma_2\delta M(\delta)$. The limits of the three sums are the three integrals of $f(x)$ in (a, b), (a, c), and (c, b) respectively; thus the theorem is established.

(5) *If* $f_1, f_2, f_3, \dots f_n$ be a finite number of limited functions, each of which is integrable in (a, b), and if $F(f_1, f_2, \dots f_n)$ be a continuous function of the n variables $f_1, f_2, \dots f_n$, then the function F is integrable in (a, b).*

For the only points of discontinuity of the function $F(x)$ are those of the functions $f_1(x), f_2(x), \dots f_n(x)$; hence the set of points of discontinuity of $F(x)$ has measure zero; and thus $F(x)$ is integrable, since it is also a limited function.

Important particular cases of the general theorem are the following:—

(*a*) If $f(x) = f_1(x) + f_2(x) + \dots + f_n(x)$, where all the functions $f_r(x)$ are integrable, then $\int_a^b f(x)\,dx = \sum_1^n \int_a^b f_r(x)\,dx$.

(*b*) If $f(x) = f_1(x) . f_2(x) \dots f_n(x)$, where all the functions $f_r(x)$ are integrable in (a, b), then $f(x)$ is also integrable in (a, b).

(*c*) If $f(x)$, $\phi(x)$ be integrable in (a, b), and $|\phi(x)|$ always exceed some fixed number A, so that $\frac{f(x)}{\phi(x)}$ is a continuous function of f and ϕ, then $\frac{f(x)}{\phi(x)}$ is integrable in (a, b).

* Du Bois Reymond, *Math. Annalen*, vol. xx, p. 123. See also W. H. Young, *Quarterly Journal of Math.* vol. xxxv, p. 190.

(6) *If two functions $f^+(x)$, $f^-(x)$ be defined as follows:—Let $f^+(x)=f(x)$ for all values of x such that $f(x)>0$, and let $f^+(x)=0$, when $f(x) \leqq 0$; let $f^-(x)=-f(x)$ for all values of x such that $f(x)<0$, and $f^-(x)=0$, when $f(x) \geqq 0$; then if $f(x)$ be integrable in (a, b), the functions $f^+(x)$, $f^-(x)$ are integrable in (a, b), and $\int_a^b f(x)\,dx = \int_a^b f^+(x)\,dx - \int_a^b f^-(x)\,dx$.*

For the fluctuation of $f^+(x)$ in any interval δ cannot exceed that of $f(x)$ in the same interval; hence, since $\Sigma\delta D(\delta)$ for $f(x)$ has the limit zero, the corresponding sum for $f^+(x)$ has the limit zero, and thus $f^+(x)$ is integrable. In a similar manner it can be shewn that $f^-(x)$ is integrable.

Since $f(x)=f^+(x)-f^-(x)$, we see from (5) (*a*) that

$$\int_a^b f(x)\,dx = \int_a^b f^+(x)\,dx - \int_a^b f^-(x)\,dx.$$

It should be observed that it is not in general true that, if $f(x)$ be integrable in (a, b), and be expressed as the sum $f_1(x)+f_2(x)$ of two limited functions, then $f_1(x)$, $f_2(x)$ are also integrable in (a, b). For it is clear that, $f(x)$ being given, we may take for $f_1(x)$ any arbitrarily defined non-integrable function, then $f_2(x)$ is also determinate and non-integrable.

(7) *If $f(x)$, $\phi(x)$ be both integrable, and be such that $|f(x)| \leqq |\phi(x)|$, for every value of x, then $\left|\int_a^b f(x)\,dx\right| \leqq \int_a^b |\phi(x)|\,dx$.*

In particular, *if $\phi(x)$ is constant and equal to P, the upper limit of $|f(x)|$ in (a, b), then $\left|\int_a^b f(x)\,dx\right| \leqq P(b-a)$.*

For $\int_a^b \{|\phi(x)|-|f(x)|\}\,dx$ is $\geqq 0$, since in every interval δ no value of $|\phi(x)|-|f(x)|$ is negative, and thus the sums of which the integral is the limit are all $\geqq 0$. Also from (2), we have $\left|\int_a^b f(x)\right|dx \leqq \int_a^b |f(x)|\,dx$, and this is $\leqq \int_a^b |\phi(x)|\,dx$. The particular case follows by assuming $\phi(x)=P$.

If U, L denote the upper and lower limits of $f(x)$ in (a, b), then

$$L(b-a) \leqq \int_a^b f(x)\,dx \leqq U(b-a).$$

For $\Sigma\delta U(\delta)$, $\Sigma\delta L(\delta)$ each lie between $U\Sigma\delta$ and $L\Sigma\delta$, or between $U(b-a)$ and $L(b-a)$; the same must hold of the common limit, which is the integral $\int_a^b f(x)\,dx$.

(8) *If $\eta_1, \eta_2, \ldots \eta_n, \ldots$ be an enumerable set of non-overlapping intervals contained in (a, b) in descending order of length, then the sum of the integrals of $f(x)$ taken through $\eta_1, \eta_2, \ldots \eta_n$, converges to a definite finite limit, as n is increased indefinitely; $f(x)$ being a function which is integrable in (a, b).*

Let us denote by S_n the sum of the integrals of $f(x)$ taken through the intervals $\eta_1, \eta_2, \dots \eta_n$. Since $\eta_1 + \eta_2 + \dots + \eta_n$ increases with n, and is always less than $b-a$, it has a definite limit as n is increased indefinitely; we can therefore choose n so great that $\eta_{n+1} + \eta_{n+2} + \dots + \eta_{n+m} < \epsilon$, for every value of m, where ϵ is an arbitrarily chosen positive number. With this value of n, we see that $|S_{n+m} - S_n| < \epsilon . P$, where P is the upper limit of $|f(x)|$ in (a, b). If η be an arbitrarily chosen positive number, we can choose ϵ such that $\epsilon < \eta/P$; thus n can be so chosen that $|S_{n+m} - S_n| < \eta$, and hence S_n has a definite limit as n is increased indefinitely.

INTEGRABLE NULL-FUNCTIONS AND EQUIVALENT INTEGRALS.

257. If $f(x)$ be integrable in (a, b), and be such that its integral in every interval contained in (a, b) is zero, then $f(x)$ is said to be an *integrable null-function*.

The necessary and sufficient condition that a limited function $f(x)$ may be an integrable null-function is that the set of points for which $|f(x)| \geqq k$, when the set is closed by the addition of its limiting points, shall be of content zero, whatever positive value k may have.

To prove that the condition stated is sufficient, let us suppose the interval (a, b) to be divided into sub-intervals by a system of sub-divisions; at any stage, let $\Sigma\delta'$ be the sum of those intervals which contain in their interiors or at their ends points at which $|f(x)| \geqq k$. The sum, of which the limit is $\int_a^b f(x)\,dx$, is in absolute value $< P\Sigma\delta' + (b - a - \Sigma\delta')k$, where P is the upper limit of $|f(x)|$ in (a, b). If the content of the closed set obtained by adding to the set for which $|f(x)| \geqq k$ its limiting points, have content zero, then $\Sigma\delta'$ has the limit zero, as the number of sub-divisions of (a, b) increases indefinitely; hence the absolute value of $\int_a^b f(x)\,dx$ is $\leqq k(b-a)$; and as k is arbitrarily small, this shews that the integral vanishes. The same argument applies to any interval (α, β) contained in (a, b). To shew that the condition is necessary, let us assume that $f(x)$ has, in every interval contained in (a, b), an integral which vanishes. At any point x_1, at which $f(x)$ is continuous, $f(x_1)$ must be zero. For let $f(x_1)$, if possible, have a positive value A; then a neighbourhood $(x_1 - h, x_1 + h)$ can be found such that at every point in it $f(x)$ lies between $A - \epsilon$ and $A + \epsilon$, where ϵ is any assigned positive number $< A$; now the integral of $f(x)$ through this interval $(x_1 - h, x_1 + h)$ is $\geqq (A - \epsilon)2h$, and thus cannot be zero, contrary to hypothesis. It is therefore impossible that $f(x_1)$ can have a positive value; and that it can have a negative value can be shewn, in a similar manner, to be also impossible; and thus $f(x)$ vanishes at every point at which it is continuous. Considering next a point at which $|f(x)| \geqq k$, or a point which is a limiting point of the set of such points, we see that the saltus at the point is $\geqq k$; for every neighbourhood of the point contains points of continuity at which $f(x)$ vanishes. The condition of integrability consequently ensures that the set of points at which $|f(x)| \geqq k$, when closed by adding the limiting points, has content zero.

The condition may also be stated in the concise form that:—

A limited function is an integrable null-function, if it vanishes at all points of a set of which the measure is equal to that of the whole interval for which the function is defined.

Two integrable functions $f(x)$, $\phi(x)$ have the same integrals in every interval contained in their domain, provided they differ from one another by an integrable null-function.

If $f(x)$ be an integrable point-wise discontinuous function, and if the function $f_1(x)$ be defined, as in § 191, by extension of that function which is defined only at the points of continuity of $f(x)$, and has at those points the same functional values as $f(x)$ itself, then $f_1(x)$, although it is in general multiple-valued at the points of discontinuity of $f(x)$, is an integrable function. It has been explained in § 255 that Riemann's definition is applicable to such a function. That $f_1(x)$ is integrable, follows from the fact that it is continuous at all the points of continuity of $f(x)$, and these form a set of points of which the measure is equal to that of the whole interval in which $f(x)$ is defined. The difference of the two functions is zero at the points of continuity of $f(x)$, and is discontinuous only at the points of discontinuity of $f(x)$, which form a set of points of zero measure. The function $f(x)-f_1(x)$ is accordingly an integrable null-function, and the two functions $f(x), f_1(x)$ have equal integrals in any interval for which $f(x)$ is defined.

It has therefore been shewn that *an integrable function $f(x)$ is equal to the sum of an integrable null-function and of the function obtained by extension of the function defined by the values of $f(x)$ at its points of continuity.*

EXAMPLES.

1. Riemann's function $f(x)=\frac{(x)}{1^2}+\frac{(2x)}{2^2}+\dots+\frac{(nx)}{n^2}+\dots$, where (x) denotes the positive or negative excess of x over the nearest integer, and $(x)=0$ when x is half-way between two integers, has been shewn in Example 2, § 190, to be point-wise discontinuous, with all its discontinuities ordinary ones, and everywhere-dense in the interval $(0, 1)$. Since all the discontinuities are ordinary ones, and the function is limited, $f(x)$ is integrable in $(0, 1)$.

2. Let $f(x)$ be defined for the interval $(0, 1)$ as follows:—If x be irrational, let $f(x)=0$; if $x=p/q$, where p/q is in its lowest terms, let $f(x)=1/q$; also let $f(0)=f(1)=0$. This function is an integrable point-wise discontinuous null-function; thus $\int_0^1 f(x)\,dx=0$.

There are only a finite number of points at which the functional value exceeds an assigned positive number.

3. Let $f(x)=0$, for all rational values of x; and $f(x)=1$, for all irrational values of x. This function is not integrable in any interval, for it is totally discontinuous.

4. Let $f(x)$ be defined* for the interval (0, 1) as follows:—For $\frac{1}{2} \leqq x \leqq 1$, let $f(x)=1$; for $\frac{1}{2^2} < x \leqq \frac{1}{2}$, let $f(x)=\frac{1}{2}$; for $\frac{1}{2^3} < x \leqq \frac{1}{2^2}$, let $f(x)=\frac{1}{2^2}$; and generally, for

$$\frac{1}{2^{n+1}} < x \leqq \frac{1}{2^n}, \text{ let } f(x)=\frac{1}{2^n}; \text{ and } f(0)=0.$$

This function is integrable, and $\int_0^x f(x)\,dx = \frac{x}{2^{m-1}} - \frac{1}{3 \cdot 2^{2m-2}}$, where x is between

$$\frac{1}{2^m} \text{ and } \frac{1}{2^{m-1}}.$$

THE FUNDAMENTAL THEOREM OF THE INTEGRAL CALCULUS.

258. The fundamental theorem of the Integral Calculus asserts that the operations of differentiation and of integration are in general inverse operations. Before we proceed to consider the conditions under which this is the case, the following theorem will be established:—

If $f(x)$ be a limited function which is integrable in the interval (a, b), then $\int_a^x f(x)\,dx$ is a continuous function of x, for the whole interval (a, b), and it is a function of limited total fluctuation in (a, b).

It has already been shewn that $\int_a^x f(x)\,dx$ exists, for any point x of the interval (a, b); denoting its value by $F(x)$, we have

$$F(x \pm h) - F(x) = \int_x^{x \pm h} f(x)\,dx;$$

hence by (7), of § 256, $|F(x \pm h) - F(x)| < Ph$, where P is the upper limit of $|f(x)|$ in (a, b). If ϵ be any arbitrarily chosen positive number, and we take $h_1 < \epsilon/P$, then for all values of h which are $\leqq h_1$, we have $|F(x \pm h) - F(x)| < \epsilon$; but this is the condition of continuity of $F(x)$ at the point x. In case x be one of the end-points of (a, b), h must be restricted to have one sign only.

To prove that $F(x)$ has limited total fluctuation in (a, b), let (a, b) be divided into n sub-intervals by the points $a, x_1, x_2, \ldots x_{n-1}, b$. The sum of the absolute differences of the values of $F(x)$ at the ends of these sub-intervals is

$$\left|\int_a^{x_1} f(x)\,dx\right| + \left|\int_{x_1}^{x_2} f(x)\,dx\right| + \ldots + \left|\int_{x_{n-1}}^{b} f(x)\,dx\right|$$

and this is, in accordance with the theorem (7) of § 256, $\leqq \int_a^b |f(x)|\,dx$; and therefore the sum is less than a fixed positive number. Since the total

* Dini, *Grundlagen*, p. 344.

variation of $F(x)$ in (a, b) is limited, it follows from the theorem of § 196, that the total fluctuation in the interval is also limited.

When $f(x)$ is integrable in (a, b), the function $\int_a^x f(x)\,dx$, which has been shewn to be continuous, and of limited total fluctuation in (a, b), is said to be the *integral function* corresponding to $f(x)$.

If $f(x)$ be any function defined in (a, b), a function $\phi(x)$ which, at every point x of the interval, possesses a differential coefficient equal to $f(x)$, is said to be an *indefinite integral* of $f(x)$.

The definition is, however, extended to cases in which $\phi'(x)$ either does not exist, or is not equal to $f(x)$, at points belonging to an exceptional set; the condition $\phi'(x) = f(x)$ being satisfied at all points not belonging to the exceptional set.

Taking the function $F(x) = \int_a^x f(x)\,dx$, as the integral function corresponding to $f(x)$, the following properties will be established:—

(A) Under certain restrictions, $F(x)$ possesses a differential coefficient which is equal to $f(x)$, and thus $F(x)$ is an indefinite integral of $f(x)$.

(B) Also it will be shewn that, if $\phi(x)$ be a function which possesses a differential coefficient $f(x)$, then $f(x)$ has in general an integral $F(x)$, in an interval (a, x), which integral differs from $\phi(x)$ by a constant only; and thus that the indefinite integral of $f(x)$ is determinate except for an additive constant.

It will appear that there are cases of exception to both theorems. When $F(x)$ is an integral function, it happens in certain cases that $F(x)$ does not possess a differential coefficient; and when $\phi(x)$ is a function which possesses a differential coefficient, it is not always the case that the latter is integrable, and when integrated yields the function $\phi(x)$ except as regards a constant.

259. *If $f(x)$ be continuous in the interval (a, b), and $F(x)$ denote the integral function $\int_a^x f(x)\,dx$, then, at every point in (a, b), $F(x)$ possesses a differential coefficient which is equal to $f(x)$.*

For since $f(x)$ is continuous, an interval $(x - h_1, x + h_1)$ can be found such that $|f(x \pm \theta h_1) - f(x)| < \epsilon$, for all proper fractional values of θ. It follows that $F(x \pm h) - F(x) \equiv \int_x^{x \pm h} f(x)\,dx$, lies between $h[f(x) + \epsilon]$ and $h[f(x) - \epsilon]$, provided $h < h_1$. Hence since $\frac{F(x \pm h) - F(x)}{h}$ lies between $f(x) + \epsilon, f(x) - \epsilon$, for $h < h_1$, it follows that $f(x)$ is the differential coefficient of $F(x)$. At the points a, b, the function $F(x)$ possesses derivatives on the right and on the left respectively, and their values are $f(a)$, $f(b)$.

If $\phi(x)$ be a function which at every point of (a, b) has a differential coefficient, which is a continuous function $f(x)$, then

$$\phi(x) - \phi(a) = \int_a^x f(x)\,dx.$$

For let $\int_a^x f(x)\,dx$ be denoted by $F(x)$, then the function $\phi(x) - F(x)$ has at every point a differential coefficient which is zero, and therefore by the theorem of § 206 the function $\phi(x) - F(x)$ is constant; it is clear that this constant must be $\phi(a)$; and thus the theorem is established. In this theorem and elsewhere, a derivative at a on the right, and a derivative at b on the left, are included in the term differential coefficient.

260. If a given limited integrable function $f(x)$ be not everywhere continuous in the integral (a, b), the proof given above is applicable to prove that, *at any point of continuity of $f(x)$, the function $\int_a^x f(x)\,dx$ has a differential coefficient equal to $f(x)$.*

At a point of ordinary discontinuity of $f(x)$, the same proof, when modified by taking only positive values of h, or only negative values of h, and using $f(x+0)$ or $f(x-0)$, in the two cases, instead of $f(x)$, will shew that $F(x)$ has at such a point derivatives on the right and on the left, and that these are $f(x+0)$, $f(x-0)$ respectively. At a point at which $f(x)$ has a discontinuity of the second kind, the proof fails altogether; at such a point therefore $F(x)$ need not possess a differential coefficient, nor definite derivatives on the right and on the left, but may have all its four derivatives $D^+F(x)$, $D_+F(x)$, $D^-F(x)$, $D_-F(x)$ of different values.

If $f(x)$ be an integrable point-wise discontinuous function, and $\psi(x)$ is the function formed by extension of the functional values of $f(x)$ at its points of continuity, as explained in § 257, we have $f(x) = \chi(x) + \psi(x)$, where $\chi(x)$ is an integrable null-function; and therefore $f(x)$ and $\psi(x)$ have the same integral function $F(x)$. The derivatives of $F(x)$ are independent of the function $\chi(x)$, and depend only upon $\psi(x)$, which is determined by the values of $f(x)$ at its points of continuity.

Since $F(x+h) - F(x) = \int_x^{x+h} \psi(x)\,dx$, and since the values of $\psi(x)$ in the interval $(x, x+h)$ all lie between $\overline{\psi(x+0)} + \epsilon_1$, and $\underline{\psi(x+0)} - \epsilon_2$, where ϵ_1, ϵ_2 converge to zero as h does so, we see that $\dfrac{F(x+h) - F(x)}{h}$ lies between $\overline{\psi(x+0)} + \epsilon_1$ and $\underline{\psi(x+0)} - \epsilon_2$, hence $D^+F(x)$, $D_+F(x)$ both lie between*

* It is stated by Schönflies, see *Bericht über die Mengenlehre*, p. 208, that the derivatives of $F(x)$ are equal to $\overline{\psi(x+0)}$, $\underline{\psi(x+0)}$, $\overline{\psi(x-0)}$, $\underline{\psi(x-0)}$. This, however, is not necessarily the case. It has been shewn by Hahn, *Monatshefte der Math. u. Physik*, vol. XVI, p. 317, that $f(x)$

$\overline{\psi(x+0)}$, $\underline{\psi(x+0)}$. By taking h negative, we see that $D^-F(x)$, $D_-F(x)$ both lie between $\overline{\psi(x-0)}$ and $\underline{\psi(x-0)}$. In case $\psi(x)$ be continuous on the right, $F(x)$ has a derivative on the right, $\psi(x+0)$; and in case $\psi(x)$ is continuous on the left, $F(x)$ has a derivative $\psi(x-0)$ on the left. It may happen that $\psi(x)$ is continuous at a point of discontinuity of $f(x)$; at such a point $F(x)$ has a differential coefficient equal to the value of $\psi(x)$. Even when $\psi(x)$ has a discontinuity of the second kind, it is possible that $F(x)$ may have a differential coefficient, or a derivative on the right or on the left, or both.

If $f(x)$ be integrable, and $F(x)$ be the corresponding integral function, any one of the four derivatives $DF(x)$ of $F(x)$ is integrable, and has $F(x)$ for its integral function.

For $DF(x)$ differs from $f(x)$ only at a point of discontinuity of $f(x)$, and at a point of discontinuity $DF(x)$ lies between the upper and lower limits of $\psi(x)$; thus $f(x) - DF(x)$ is an integrable null-function. Therefore

$$\int_a^x \psi(x)\,dx = \int_a^x D^+F(x)\,dx = \int_a^x D_+F(x)\,dx = \int_a^x D^-F(x)\,dx$$
$$= \int_a^x D_-F(x)\,dx = \int_a^x f(x)\,dx = F(x).$$

It has been shewn that the integral function of an integrable point-wise discontinuous function has a differential coefficient at the everywhere-dense set of points of continuity of the discontinuous function; there may however also be an everywhere-dense set of points at which this continuous function does not possess a differential coefficient.

261. It has been shewn that, if the continuous function $\phi(x)$ possesses everywhere a differential coefficient $f(x)$ which is everywhere a continuous function, then

$$\phi(x) - \phi(a) = \int_a^x f(x)\,dx = F(x).$$

This is a particular case of the following more general theorem:—

If $\phi(x)$ be a function continuous in the interval (a, b), and if one of its four derivatives $D^+\phi(x)$, $D_+\phi(x)$, $D^-\phi(x)$, $D_-\phi(x)$ be a limited integrable function in (a, b), then each of the other three derivatives is also limited and integrable in (a, b), and $\phi(x) - \phi(a)$ is the integral of any one of the four derivatives through the interval (a, x).

If (a, x) be divided into a number of parts (a, x_1), (x_1, x_2), ... (x_{n-1}, x), it

may be so chosen that the corresponding integral function has, at a particular point, derivatives on the right having arbitrarily given values lying between, or equal to, the values of $\overline{\psi(x+0)}$, $\underline{\psi(x+0)}$ at the point; and in particular that $f(x)$ may be so constructed as to have, at the point, a definite derivative which has an assigned value between the two limits.

has been shewn in § 217, that $\frac{\phi(x_r)-\phi(x_{r-1})}{x_r - x_{r-1}}$ lies between the upper and lower limits of any one of the four derivatives $D\phi(x)$ in the interval (x_{r-1}, x_r). It follows that $\phi(x)-\phi(a)$ lies between the two sums

$$(x_1-a)\,U(a, x_1, D)+(x_2-x_1)\,U(x_1, x_2, D)+\ldots+(b-x_{n-1})\,U(x_{n-1}, b, D),$$
$$(x_1-a)\,L(a, x_1, D)+(x_2-x_1)\,L(x_1, x_2, D)+\ldots+(b-x_{n-1})\,L(x_{n-1}, b, D),$$

where $U(x_{r-1}, x_r, D)$, $L(x_{r-1}, x_r, D)$ are the upper and lower limits of $D\phi(x)$ in the interval (x_{r-1}, x_r); and it is known that these are the same for all four derivatives. The limits of the above sums, when the intervals are diminished indefinitely, so that the greatest of them converges to zero, are the upper and lower integrals of any one of the four functions $D\phi(x)$. If it be known that any one of these derivatives is integrable in (a, x), then the upper and lower integrals are equal, and the other three are also integrable, the common value of the integral being $\phi(x)-\phi(a)$. Thus

$$\phi(x)-\phi(a)=\int_a^x D^+\phi(x)\,dx=\int_a^x D_+\phi(x)\,dx=\int D^-\phi(x)\,dx=\int D_-\phi(x)\,dx.$$

It should be observed that, as has been shewn in § 219, the four derivatives are all equal to one another at a point at which one of them is a continuous function; and thus at such a point there is a differential coefficient. If one of the derivatives be integrable, there is therefore a set of points of measure equal to that of the interval (a, b), at which all four derivatives have equal values, and at which therefore a differential coefficient exists.

262. In case $D\phi(x)$ be a limited function which is not integrable, the above proof shews that $\phi(x)-\phi(a)$ lies between the upper and lower integrals in (a, x) of any one of the four functions $D\phi(x)$. This includes the case in which $\phi(x)$ has a differential coefficient which is limited but not integrable; in that case $\phi(x)-\phi(a)$ lies between $\overline{\int_a^x}\phi'(x)\,dx$ and $\underline{\int_a^x}\phi'(x)\,dx$.

Since $\overline{\int_x^{x+h}}\phi'(x)\,dx$ is in absolute value less than $h.U$, where U is the upper limit of $\phi'(x)$ in (a, b), it follows as in § 258, that $\overline{\int_a^x}\phi'(x)\,dx$ is a continuous function of x; similarly it may be seen that $\underline{\int_a^x}\phi'(x)\,dx$ is a continuous function of x. At a point of continuity of $\phi'(x)$, both $\overline{\int_a^x}\phi'(x)\,dx$, $\underline{\int_a^x}\phi'(x)\,dx$ have the differential coefficient $\phi'(x)$, as may be seen by a process precisely similar to that in § 259. Thus the upper and lower integrals of $\phi'(x)$ possess properties similar to those of the integral of $\phi'(x)$ when it exists.

The function $D\phi(x)$ when not integrable, may be a non-integrable point-wise discontinuous function, or it may be totally discontinuous.

If $f(x)$ be any non-integrable limited function, the following theorems may be established by proofs similar to those in § 258 and § 259:—

The upper and lower integrals $\overline{\int_a^x} f(x)\,dx$, $\underline{\int_a^x} f(x)\,dx$ *are continuous, and of limited total fluctuation in* (a, b).

At any point of (a, b) *at which* $f(x)$ *is continuous, the upper and lower integrals* $\overline{\int_a^x} f(x)\,dx$, $\underline{\int_a^x} f(x)\,dx$ *each possess a differential coefficient which is equal to* $f(x)$.

263. An important general class of continuous functions for which the four derivatives are not integrable, even when a differential coefficient exists, or when derivatives on the right and on the left always exist, is the class of everywhere-oscillating functions. Those functions which become everywhere-oscillating functions when a linear function is added have the same property.

If a derivative $DF(x)$ be such that in every interval it has no finite upper limit or no finite lower limit, it is certainly not integrable; it is therefore only necessary to consider an interval in which the function $DF(x)$ is limited. Let (a, x) be such an interval, and let us suppose that $F(x) - F(a)$ is not zero.

In every interval (x_{r-1}, x_r) contained in (a, x), $U(x_{r-1}, x_r, D)$ the upper limit of $DF(x)$ is positive, and $L(x_{r-1}, x_r, D)$ the lower limit of $DF(x)$ is negative; thus the two sums

$$(x_1 - a)\,U(a, x_1, D) + (x_2 - x_1)\,U(x_1, x_2, D) + \dots + (b - x_{n-1})\,U(x_{n-1}, b, D),$$

$$(x_1 - a)\,L(a, x_1, D) + (x_2 - x_1)\,L(x_1, x_2, D) + \dots + (b - x_{n-1})\,L(x_{n-1}, b, D),$$

are such that the first is essentially positive, and the second essentially negative, the non-vanishing number $F(x) - F(a)$ lying between them.

It follows that the limits of these two sums, as the number of subdivisions of (a, x) is increased indefinitely, must be different from one another, since they cannot have zero as their common value; thus

$$\overline{\int_a^x} DF(x)\,dx,\quad \underline{\int_a^x} DF(x)\,dx$$

are distinct from one another.

It has thus been proved that *a continuous function which is everywhere-oscillating in* (a, b) *cannot have a derivative which is integrable in* (a, b), *even if it have everywhere a differential coefficient, or definite derivatives on the right and on the left.*

The function $DF(x)$, or $f(x)$, in the case of such function, may be a point-wise discontinuous function such that the measure of the set of points of discontinuity is greater than zero, or it may be a totally discontinuous function.

A continuous monotone function, which is not reducible to a function with an infinite number of oscillations by the addition of a linear function, has at every point definite derivatives on the right and on the left, each of which is either continuous or is an integrable point-wise discontinuous function, since either derivative has only ordinary discontinuities. Thus such a function has integrable derivatives, provided these derivatives are limited in the interval.

In case the continuous function $F(x)$ have a differential coefficient, or a derivative which is not everywhere finite, or is not limited in the interval, this derivative is not integrable in the sense in which we have hitherto defined integration. This case will be considered in connection with the theory of improper integrals.

264. The preceding investigations provide answers to the questions which arise as regards the validity of the two propositions (A) and (B) of § 258, which together constitute the fundamental theorem of the Integral Calculus asserting that the operations of differentiation and of integration are in general reversible. The definition of a definite integral has hitherto been restricted to that of Riemann, and is applicable to limited functions only. The extensions of that definition to the case of unlimited functions, which will be considered later, and also a more general definition of integration due to Lebesgue, of which an account will also be given, will lead to corresponding extensions of the scope of the fundamental theorem.

As regards the theorem (A), that the integral function $F(x) \equiv \int_a^x f(x)\,dx$ of a limited integrable function possesses a differential coefficient equal, at a point x of (a, b), to $f(x)$, it has been shewn that the theorem holds without restriction in case $f(x)$ is a continuous function; but that, if $f(x)$ be not continuous, the theorem still holds as regards every point of continuity of $f(x)$. It follows that the points of (a, b) at which $F(x)$ either possesses no differential coefficient, or possesses one which is not equal to $f(x)$, form a set of zero measure, which may however be everywhere-dense in (a, b).

The theorem (B) that, if $\phi(x)$ possess a differential coefficient $f(x)$, then the corresponding integral function $F(x) \equiv \int_a^x f(x)\,dx$ differs from $\phi(x)$ only by a constant, holds if $f(x)$ be a continuous function, and more generally, if $f(x)$ be limited and integrable. In case $\phi(x)$ do not at all points possess a differential coefficient, the more general theorem is applicable that, if any one of the four derivatives of $\phi(x)$ be limited and integrable, then the integral function corresponding to that derivative differs from $\phi(x)$ by a constant only. The theorem fails either in case $\phi(x)$ be not a function with limited derivatives, or in case it be a function with limited derivatives, but those derivatives do not satisfy Riemann's condition of integrability.

The problem of the determination of a continuous function which shall have a given function $f(x)$ for its differential coefficient, at every point at which $f(x)$ is continuous, may be here considered in the case in which $f(x)$ is restricted to be limited in the interval (a, b) for which it is defined. This problem is regarded as having a determinate solution provided functions exist which satisfy the condition, and further provided any two such functions differ from one another by a constant only, that constant having one and the same value for the whole interval (a, b). In the first place, the problem cannot be determinate unless $f(x)$ be integrable; for either of the two functions $\overline{\int_a^x} f(x)\,dx$, $\underline{\int_a^x} f(x)\,dx$ satisfies the condition of the problem, and these functions do not differ from one another by a constant, as they both vanish at the point a, and are elsewhere unequal. Next, if $f(x)$ be integrable, the function $\int_a^x f(x)\,dx$ satisfies the condition of the problem, but the solution is not necessarily determinate. In case however the points of discontinuity of the integrable function $f(x)$ form an enumerable set, the theorem of § 206 shews that the solution is determinate; for any two functions which have equal finite differential coefficients at all points of (a, b) except those of an enumerable set, differ from one another by a constant. In this case $\int_a^x f(x)\,dx + C$ is the function required. When the points of discontinuity of the integrable function $f(x)$ form an unenumerable set, although that set must have zero measure, the problem has not a determinate solution. For, although $\int_a^x f(x)\,dx$ is a function which has the required property, another solution is obtained by adding to it any continuous function which has all the intervals complementary to the perfect component of the unenumerable set as lines of invariability; that such functions exist has been established in § 208. There exists however only one function, viz. $\int_a^x f(x)\,dx$, with limited derivatives, which satisfies the condition of the problem; for it has been shewn in § 224, that any two functions which have limited derivatives, one of which derivatives is prescribed at all points not belonging to a certain set of measure zero, differ from one another by a constant.

Similar remarks apply to the more general problem of the determination of a function which shall have one of its four derivatives, say the upper one on the right, equal to a given function $f(x)$ at every point of continuity of $f(x)$. This problem has a solution whenever $f(x)$ is limited; in virtue of the theorem of § 206, the solution is determinate when $f(x)$ is integrable, and the points of discontinuity of $f(x)$ form an enumerable set. When $f(x)$ is integrable, and the set of points of discontinuity is unenumerable, there exists, in virtue of the theorem of § 224, only one solution for which the derivatives are limited. As before, if the restriction, that the required function is to have limited derivatives be not imposed, the solution of the problem is indeterminate.

EXAMPLE.

Let G be a perfect non-dense set of points in the interval (a, b), and such that its content is greater than zero. Let (α, β) be an interval complementary to the set G, and let $\phi(x, \alpha)=(x-\alpha)^2 \sin\frac{1}{x-\alpha}$, and therefore $\phi'(x, \alpha)=2(x-\alpha)\sin\frac{1}{x-\alpha}-\cos\frac{1}{x-\alpha}$. The function $\phi'(x, \alpha)$ vanishes at an infinite number of points in (α, β); let $\alpha+\gamma$ be the greatest value of x which does not exceed $\frac{1}{2}(\alpha+\beta)$, for which $\phi'(x, \alpha)$ vanishes. Let $F(x)=0$ at every point of G, and in each interval (α, β) complementary to G, let $F(x)=\phi(x, \alpha)$, for values of x such that $\alpha \leqq x \leqq \alpha+\gamma$; let $F(x)=\phi(\alpha+\gamma, \alpha)$ for values of x such that $\alpha+\gamma \leqq x \leqq \beta-\gamma$; and let $F(x)=-\phi(x, \gamma)$ for $\beta-\gamma \leqq x \leqq b$. The function $F(x)$ is continuous, and has everywhere a finite differential coefficient which is limited in the interval (a, b). It is easily seen that $F'(x)$ vanishes at every point of G. The function $F'(x)$ has a discontinuity of measure 2 at each point of the set G which is not of zero content, and therefore $F'(x)$ is not an integrable function. This example was given* by Volterra, as the first known example of a continuous function possessing a non-integrable limited differential coefficient.

FUNCTIONS WHICH ARE LINEAR IN EACH INTERVAL OF A SET.

265. The existence of continuous functions which are linear in each interval of an everywhere-dense set of intervals has been already referred to in § 213. It has been shewn in § 208 how a function $f(x)$ can be constructed which is continuous, and has as lines of invariability the intervals complementary to a non-dense perfect set of points. It is clear that the integral function $\int_a^x f(x)\,dx$ is linear in each of the intervals, and being also continuous, it is a function of the type referred to. A more general function which is continuous, and is linear in each interval of the set, may be obtained by adding to $\int_a^x f(x)\,dx$ any continuous function for which the intervals of the set are lines of invariability.

* *Giorn. di Battaglini*, vol. XIX, 1881.

MEAN VALUE THEOREMS.

266. It is frequently of importance to be able to assign upper and lower limits between which the value of a definite integral lies, in cases where the exact determination of the value of the definite integral is not required. Such estimates of the value of a definite integral may frequently be made by means of theorems known as mean value theorems; the most important of these will be here given.

If $f(x)$ be a limited integrable function defined for the interval (a, b), it is clear from the definition of the integral as the limit of a sum, that if U, L denote the upper and lower limits of $f(x)$ in the interval (a, b), then

$$L(b-a) \leqq \int_a^b f(x)\,dx \leqq U(b-a);$$

it follows that
$$\int_a^b f(x)\,dx = (b-a)M,$$

where M is some number which satisfies the condition $U \geqq M \geqq L$.

In case $f(x)$ is a continuous function, there must be some value or values of x in (a, b) for which $f(x) = M$; if then such a point be denoted by $a + \theta(b-a)$, we obtain the following theorem:—

If $f(x)$ be continuous in the interval (a, b), then

$$\int_a^b f(x)\,dx = (b-a) f\{a + \theta(b-a)\},$$

where θ is some number such that $0 \leqq \theta \leqq 1$.

Next, let $f(x)$, $\phi(x)$ be integrable functions defined for the interval (a, b), the function $\phi(x)$ being positive, or zero, for the whole interval; we then have immediately from the definition of the integral

$$\int_a^b f(x)\,\phi(x)\,dx,$$

$$L\int_a^b \phi(x)\,dx \leqq \int_a^b f(x)\,\phi(x)\,dx \leqq U\int_a^b \phi(x)\,dx,$$

where, as before, U, L are the upper and lower limits of $f(x)$ in (a, b).

It follows at once, that

$$\int_a^b f(x)\,\phi(x)\,dx = M_1 \int_a^b \phi(x)\,dx,$$

where M_1 is some number such that

$$U \geqq M_1 \geqq L.$$

In case $f(x)$ be a continuous function, we obtain the following theorem:—

If $f(x)$ be continuous in (a, b), and $\phi(x)$ be a limited integrable function which has the same sign throughout (a, b), except where it may be zero, then

$$\int_a^b f(x)\,\phi(x)\,dx = f\{a+\theta_1(b-a)\}\int_a^b \phi(x)\,dx,$$

where θ_1 is some number such that $0 \leqq \theta_1 \leqq 1$.

This theorem, including also the more general case in which $f(x)$ is not continuous, is known as the First Mean Value Theorem.

An extension to the case in which $\phi(x)$ is not necessarily everywhere of the same sign in the interval (a, b), may be obtained by applying the theorem to $\phi(x)+C$, where C is so chosen that this latter function is of invariable sign in the interval.

267. *If the limited function $f(x)$ be everywhere positive in the interval (a, b), and never increase as x increases from a to b, and is consequently integrable, and if $\phi(x)$ be limited and integrable in (a, b), then*

$$\int_a^b f(x)\,\phi(x)\,dx = f(a)\int_a^\xi \phi(x)\,dx,$$

where ξ is some number such that

$$a \leqq \xi \leqq b.$$

Also if $f(x)$ be everywhere positive, and never diminish as x increases from a to b, then

$$\int_a^b f(x)\,\phi(x)\,dx = f(b)\int_\xi^b \phi(x)\,dx,$$

where ξ is some number such that $a \leqq \xi \leqq b$.

This theorem was first given* by Bonnet, and applied by him to the theory of Fourier's series.

Another form of the theorem was obtained by Weierstrass, and also by† Du Bois Reymond, and is generally known as the Second Mean Value Theorem. This is as follows:—

If $f(x)$ be monotone and therefore integrable in (a, b), and if $\phi(x)$ be limited and integrable in the same interval, then

$$\int_a^b f(x)\,\phi(x)\,dx = f(a)\int_a^\xi \phi(x)\,dx + f(b)\int_\xi^b \phi(x)\,dx,$$

where ξ is some point in the interval (a, b).

The theorem in this form is deducible immediately from Bonnet's theorem, by writing in that theorem $f(x)-f(b)$, or $f(x)-f(a)$, in the two cases,

* See *Mém. Acad. Belg.* vol. XXIII (1850), p. 8; also *Liouville's Journal*, vol. XIV (1849), p. 249.

† *Crelle's Journal*, vol. LXIX (1869), p. 81.

instead of $f(x)$; Bonnet's theorem is however not* immediately deducible from the second theorem.

It is clear that, since the value of

$$\int_a^b f(x)\,\phi(x)\,dx$$

is unaltered by changing the values of $f(a)$ and $f(b)$, we may in the above statement instead of $f(a), f(b)$ take any two numbers† A, B which are such that the function $\psi(x)$ defined by

$$\psi(a) = A,\ \ \psi(b) = B,\ \ \psi(x) = f(x),$$

for $a < x < b$, is monotone. We have thus the following generalized form of the theorem:—

$$\int_a^b f(x)\,\phi(x)\,dx = A\int_a^\xi \phi(x)\,dx + B\int_\xi^b \phi(x)\,dx,$$

where $$A \leqq f(a+0),\ \ B \geqq f(b-0),$$

if $f(x)$ *never decreases; or*

$$A \geqq f(a+0),\ \ B \leqq f(b-0),$$

if $f(x)$ *never increases.*

The value of ξ depends in general upon the chosen values of A and B. In this generalized form, the theorem includes Bonnet's theorem as a particular case. For we may take $A = 0$, $B = f(b)$, if $f(x)$ is positive and never decreases, as x increases from a to b; or $A = f(a)$, $B = 0$, in case $f(x)$ is positive and never diminishes, as x increases from a to b.

Various proofs‡ of the theorem of Weierstrass and Du Bois Reymond§ have been given. In these proofs the function $\phi(x)$ is usually restricted to change its sign only a finite number of times; but a proof free from that restriction was given by Du Bois Reymond. A proof has been given by Pringsheim‖ in which $\phi(x)$ is not restricted to be a limited function, but may be any function such that it has an absolutely convergent integral, or in certain cases it may have an integral which does not converge absolutely.

The following proof, in which $\phi(x)$ is restricted to be a limited function, is due to Hölder¶.

* This was pointed out by Pringsheim, *Münchener Berichte*, vol. XXX (1900), where an account of various proofs of the theorem is given.

† Du Bois Reymond, *Schlömilch's Zeitschrift*, vol. XX (1875); *Hist. Lit. Abtg.* p. 126.

‡ For example, by Hankel, *Schlömilch's Zeitschrift*, vol. XIV (1869); by Meyer, *Math. Annalen*, vol. VI (1872); by C. Neumann, *Kreis- Kugel- und Cylinderfunctionen*, Leipzig, 1881, p. 28.

§ *Crelle's Journal*, vol. LXXIX (1875), p. 42. See also Kronecker's *Vorlesungen*, vol. I.

‖ *Loc. cit.*

¶ *Göttinger Anzeigen*, 1894, p. 519. Another proof has been given by Netto, *Schlömilch's Zeitschrift*, vol. XL (1895). See also Kowalewski, *Math. Ann.* vol. LX, 1905.

Let a, b be denoted by a_0, a_n, and let $a_1, a_2, a_3, \dots a_{n-1}$ be points in (a, b) in order from left to right. We consider the sum

$$\sum_{\nu=0}^{n-1}(a_{\nu+1}-a_\nu)f(c_\nu)\,\phi(c_\nu), \quad\text{.........................(1)}$$

where c_ν is any point in the interval $(a_\nu, a_{\nu+1})$; the limit of this sum defines the integral

$$\int_a^b f(x)\,\phi(x)\,dx.$$

This sum may be transformed into

$$\sum_{\nu=0}^{n-2}\{[f(c_\nu)-f(c_{\nu+1})]\sum_{\mu=0}^{\nu}(a_{\mu+1}-a_\mu)\,\phi(c_\mu)\}$$
$$+f(c_{n-1})\sum_{\mu=0}^{n-1}(a_{\mu+1}-a_\mu)\,\phi(c_\mu). \quad\text{......(2)}$$

By increasing the number of points in the interval, we may obtain a convergent sequence of sub-divisions, then

$$f(c_{n-1})\sum_{\mu=0}^{n-1}(a_{\mu+1}-a_\mu)\,\phi(c_\mu)$$

has as limit

$$f(b-0)\int_a^b \phi(x)\,dx.$$

We have now to examine the sum

$$\sum_{\nu=0}^{n-2}\{[f(c_\nu)-f(c_{\nu+1})]\sum_{\mu=0}^{\nu}(a_{\mu+1}-a_\mu)\,\phi(c_\mu)\}.$$

Since $f(x)$ is monotone, this is equal to

$$M\sum_{\nu=0}^{n-2}[f(c_\nu)-f(c_{\nu+1})],$$

where M is between the greatest and least of the numbers

$$\sum_{\mu=0}^{\nu}(a_{\mu+1}-a_\mu)\,\phi(c_\mu).$$

We have now, from (1) and (2),

$$\sum_{\nu=0}^{n-1}(a_{\nu+1}-a_\nu)\,f(c_\nu)\,\phi(c_\nu)$$
$$-f(c_{n-1})\sum_{\mu=0}^{n-1}(a_{\mu+1}-a_\mu)\,\phi(c_\mu)=M\{f(c_0)-f(c_{n-1})\}. \quad\text{...(3)}$$

In order to estimate the value of M, we have

$$\sum_{\mu=0}^{\nu}(a_{\mu+1}-a_\mu)\,\phi(c_\mu)-\int_a^{a_{\nu+1}}\phi(x)\,dx=\sum_{\mu=0}^{\nu}\int_{a_\mu}^{a_{\mu+1}}\{\phi(c_\mu)-\phi(x)\}\,dx;$$

the absolute value of this difference is at most

$$\sum_{\mu=0}^{\nu}(a_{\mu+1}-a_\mu)\,\Delta_\mu \leqq \sum_{\mu=0}^{n-1}(a_{\mu+1}-a_\mu)\,\Delta_\mu,$$

where Δ_μ is the difference between the upper and lower limits of $\phi(x)$ in the interval $(a_\mu, a_{\mu+1})$. We have therefore,

$$\sum_{\mu=0}^{\nu}(a_{\mu+1}-a_\mu)\phi(c_\mu)=\int_a^{a_{\nu+1}}\phi(x)\,dx+\theta\sum_{\mu=0}^{n-1}(a_{\mu+1}-a_\mu)\Delta_\mu,$$

where $$-1 \leqq \theta \leqq 1.$$

If G be the greatest, and H the least value of the continuous function

$$\int_a^\xi \phi(x)\,dx,$$

of ξ, in the interval (a, b), we have now, from (3),

$$H-\sum_{\mu=0}^{n-1}(a_{\mu+1}-a_\mu)\Delta_\mu \leqq M \leqq G+\sum_{\mu=0}^{n-1}(a_{\mu+1}-a_\mu)\Delta_\mu.$$

We now obtain from (3), the inequalities

$$\begin{aligned}\{H-\sum_{\mu=0}^{n-1}(a_{\mu+1}-a_\mu)\Delta_\mu\}&[f(c_0)-f(c_{n-1})]\\ &\leqq \sum_{\nu=0}^{n-1}(a_{\nu+1}-a_\nu)f(c_\nu)\phi(c_\nu)-f(c_{n-1})\sum_{\mu=0}^{n-1}(a_{\mu+1}-a_\mu)\phi(c_\mu)\\ &\leqq \{G+\sum_{\mu=0}^{n-1}(a_{\mu+1}-a_\mu)\Delta_\mu\}[f(c_0)-f(c_{n-1})],\end{aligned}$$

provided that $f(x)$ does not decrease as x increases in the interval (a, b); in the other case, the signs $<$ and $>$ must be interchanged.

We obtain now, by proceeding to the limit,

$$\begin{aligned}H\{f(a+0)-f(b-0)\} &\leqq \int_a^b f(x)\phi(x)\,dx\\ &\quad -f(b-0)\int_a^b \phi(x)\,dx \leqq G\{f(a+0)-f(b-0)\};\end{aligned}$$

hence $$\int_a^b f(x)\phi(x)\,dx-f(b-0)\int_a^b\phi(x)\,dx=M'\{f(a+0)-f(b-0)\},$$

where M' is between H and G. There must be a value of X in the interval (a, b), such that

$$M'=\int_a^X \phi(x)\,dx;$$

we therefore obtain the equation

$$\int_a^b f(x)\phi(x)\,dx=f(a+0)\int_a^X\phi(x)\,dx+f(b-0)\int_X^b\phi(x)\,dx.$$

268. The theorem thus obtained is not identical with the theorem of Du Bois Reymond and Weierstrass in its original form, since $f(a+0)$, $f(b-0)$ take the places of $f(a)$, $f(b)$; it is however a particular case of the generalized theorem which includes the original one. We therefore proceed to deduce the generalized theorem from the special form obtained in Hölder's proof.

Let the monotone integrable function $F(x)$ be defined by the conditions

$$F(x) = A, \text{ for } a \leq x \leq a + \epsilon,$$
$$F(x) = f(x), \text{ for } a + \epsilon < x < b - \epsilon,$$
$$F(x) = B, \text{ for } b - \epsilon \leq x \leq b;$$

we can then apply the foregoing result to $F(x)$ instead of $f(x)$; it is clear that

$$F(a+0) = A, \quad F(b-0) = B.$$

We have now

$$\int_a^b F(x)\,\phi(x)\,dx = A\int_a^\xi \phi(x)\,dx + B\int_\xi^b \phi(x)\,dx;$$

but $$\int_a^b f(x)\,\phi(x)\,dx - \int_a^b F(x)\,\phi(x)\,dx$$
$$= \int_a^{a+\epsilon} [f(x) - A]\,\phi(x)\,dx + \int_{b-\epsilon}^b [f(x) - B]\,\phi(x)\,dx.$$

The absolute value of the expression on the right-hand side is less than $\epsilon(P+Q)$, where P, Q are the upper limits of

$$|\{f(x) - A\}\,\phi(x)|,$$
$$|\{f(x) - B\}\,\phi(x)|$$

in the interval (a, b). Now ϵ is arbitrarily small, hence

$$\int_a^b f(x)\,\phi(x)\,dx - A\int_a^\xi \phi(x)\,dx - B\int_\xi^b \phi(x)\,dx,$$

where ξ depends on ϵ, is $< (P+Q)\,\epsilon$; hence

$$\int_a^b f(x)\,\phi(x)\,dx = A\int_a^{\xi_1} \phi(x)\,dx + B\int_{\xi_1}^b \phi(x)\,dx,$$

where ξ_1 is the limit of ξ, when ϵ converges to zero. The condition that $F(x)$ is monotone is satisfied, if

$$A \leq f(a+0), \quad B \geq f(b-0),$$

when $f(x)$ never decreases as x increases from a to b; and if

$$A \geq f(a+0), \quad B \leq f(b-0),$$

in case $f(x)$ never increases.

It will be observed that the sole conditions which are attached to the functions $f(x)$, $\phi(x)$, are that they be both limited, and integrable, and that $f(x)$ be monotone; both functions may have any set of discontinuities which is consistent with integrability, and no assumption is made as to their differentiability. The proof of the theorem given by Weierstrass depended upon an integration by parts, in which the existence of a differential coefficient of one of the functions is a necessary assumption; this assumption would place a restriction on the validity of the theorem which would unduly limit its applicability to investigations such as those connected with the theory of Fourier's series.

269. If the function $f(x)$ be not monotone, but be such that the interval (a, b) can be divided into a finite number of portions in each one of which $f(x)$ is monotone, the second mean value theorem may be applied to the integral taken through each of these portions separately; we then have

$$\int_a^b f(x)\phi(x)\,dx = f(a+0)\int_a^{x_1}\phi(x)\,dx + f(a_2-0)\int_{x_1}^{a_2}\phi(x)\,dx$$
$$+ f(a_2+0)\int_{a_2}^{x_2}\phi(x)\,dx + f(a_3-0)\int_{x_2}^{a_3}\phi(x)\,dx$$
$$+ \ldots\ldots\ldots\ldots\ldots$$

where (a, a_2), (a_2, a_3) ... are the intervals in each of which $f(x)$ is monotone.

When $f(x)$ is a function with an infinite number of oscillations, and is of the first species, a function $f(x) - lx$ can be found which is monotone in (a, b); we then have

$$\int_a^b f(x)\phi(x)\,dx = f(a)\int_a^{\xi}\phi(x)\,dx + f(b)\int_{\xi}^{b}\phi(x)\,dx$$
$$- la\int_a^{\xi}\phi(x)\,dx - lb\int_{\xi}^{b}\phi(x)\,dx$$
$$+ l\int_a^b x\phi(x)\,dx, \quad \text{where } a \leqq \xi \leqq b.$$

Again, $$\int_a^b x\phi(x)\,dx = a\int_a^{\xi'}\phi(x)\,dx + b\int_{\xi'}^{b}\phi(x)\,dx,$$

where $a \leqq \xi' \leqq b$; we thus find that

$$\int_a^b f(x)\phi(x)\,dx = f(a)\int_a^{\xi}\phi(x)\,dx$$
$$+ f(b)\int_{\xi}^{b}\phi(x)\,dx - l(b-a)\int_{\xi}^{\xi'}\phi(x)\,dx,$$

where l is, in accordance with § 214, any number which does not lie between the upper and lower limits of the derivatives of $f(x)$; and the values of ξ, ξ' will in general depend upon the value of l chosen.

IMPROPER INTEGRALS.

270. The definition of a definite integral becomes nugatory if, in the interval (a, b), there exists any sub-interval in which the upper limit or the lower limit of the function is indefinitely great. In such a case the function is not limited, and the sums whose limits are the upper and lower integrals become in one case or in both cases indefinitely great; and thus the function is not integrable in the sense defined.

Let us suppose that a point c, where $a < c < b$, is such that in its arbitrarily small neighbourhood the function has no upper or no lower limit, and let us suppose further that c is the only point of this kind, so that the function $f(x)$ is integrable in any sub-interval of (a, b) which does not contain c in its interior or at an end. The two integrals

$$\int_a^{c-\epsilon} f(x)\,dx, \quad \int_{c+\epsilon'}^b f(x)\,dx$$

both exist, whatever sufficiently small positive values be assigned to ϵ, ϵ'.

It may happen that, as ϵ, ϵ' are diminished independently so as to converge in each case to the limit zero, the two integrals also converge to definite limits; if this be the case we define the sum

$$\lim_{\epsilon=0} \int_a^{c-\epsilon} f(x)\,dx + \lim_{\epsilon'=0} \int_{c+\epsilon'}^b f(x)\,dx$$

to be the improper integral of $f(x)$ in the interval (a, b), and we denote this improper integral by

$$\int_a^b f(x)\,dx,$$

using the same notation as in the case in which $f(x)$ is integrable in (a, b).

The condition that

$$\lim_{\epsilon=0} \int_a^{c-\epsilon} f(x)\,dx$$

should exist is that, corresponding to each arbitrarily small number δ which may be chosen, a number ϵ_1 can be found such that

$$\left| \int_{c-\theta\epsilon_1}^{c-\epsilon_1} f(x)\,dx \right| < \delta,$$

whatever value θ may have, subject to the condition $0 < \theta < 1$.

A similar condition must be satisfied in order that

$$\lim_{\epsilon'=0} \int_{c+\epsilon'}^b f(x)\,dx$$

may exist.

It may happen that, although the two limits

$$\int_a^{c-\epsilon} f(x)\,dx, \quad \int_{c+\epsilon'}^b f(x)\,dx$$

do not exist, yet if we take $\epsilon' = \epsilon$, the sum

$$\int_a^{c-\epsilon} f(x)\,dx + \int_{c+\epsilon}^b f(x)\,dx$$

may have a definite limit; when that is the case, this limit defines Cauchy's ***principal value*** of the integral of $f(x)$ in (a, b).

It thus appears that a principal value may exist when the function possesses neither an integral nor an improper integral in the interval (a, b).

In case the point a itself be a point of infinite discontinuity, then the limit

$$\int_{a+\epsilon}^{b} f(x)\,dx,$$

for $\epsilon = 0$, when it exists, is defined to be the improper integral

$$\int_{a}^{b} f(x)\,dx$$

of $f(x)$ in the interval (a, b).

A similar definition applies in case the point b is a point of infinite discontinuity of the function.

If in the interval (a, b) there are two points of infinite discontinuity c_1, c_2 where $(a < c_1 < c_2 < b)$, then take any point c between c_1 and c_2. In case the four improper integrals

$$\int_{a}^{c_1} f(x)\,dx,\quad \int_{c_1}^{c} f(x)\,dx,\quad \int_{c}^{c_2} f(x)\,dx,\quad \int_{c_2}^{b} f(x)\,dx$$

all exist, their sum is defined to be the improper integral of $f(x)$ in (a, b), and is denoted by

$$\int_{a}^{b} f(x)\,dx;$$

and it is clearly independent of the value of c. The definition in case one or both of the points c_1, c_2 are end-points of the interval (a, b) is of the same character. If $c_1 = a$, $c_2 = b$, then if the two improper integrals

$$\int_{a}^{c} f(x)\,dx,\quad \int_{c}^{b} f(x)\,dx$$

exist, their sum defines the improper integral

$$\int_{a}^{b} f(x)\,dx.$$

The definition of an improper integral is now immediately extensible to the case in which there are any finite number of points of infinite discontinuity in the interval. If these be $c_1, c_2, c_3 \ldots c_n$ taken in order from left to right, then if the improper integrals

$$\int_{a}^{c_1} f(x)\,dx,\quad \int_{c_1}^{c_2} f(x)\,dx, \ldots \int_{c_n}^{b} f(x)\,dx$$

all exist, their sum is defined to be the improper integral

$$\int_{a}^{b} f(x)\,dx.$$

271. The definition of an improper integral was extended by Du Bois Reymond and by Dini to the case of a function with an indefinitely great number of points of infinite discontinuity forming a set of the first species.

The definition has, however, been further extended by Harnack to the more comprehensive case in which the set of points of infinite discontinuity is any set of zero content. The set is closed, since any limiting point of points of infinite discontinuity is also such a point. It is sometimes convenient to include in such a set, points at which the functional value is regarded as indefinitely great. Unless the upper or lower limit of the function for an arbitrarily small neighbourhood of such a point is also, when the functional value at the point is disregarded, indefinitely great, such a discontinuity is a removable infinite discontinuity.

The general definition of an improper integral is obtained by extending the principle which has been applied to the case in which the number of points of infinite discontinuity is finite, namely that the neighbourhoods of all the infinities are excluded in taking the integral. If the integral is to be considered as in any sense belonging to the whole interval, the sum of the excluded parts of the integral should have the limit zero; and thus the case in which the content of the closed set of infinite discontinuities is zero indicates the extreme extension which can fairly be given to the meaning of an improper integral through a given interval.

Let the points of infinite discontinuity of the function $f(x)$, defined for the interval (a, b), form a non-dense closed set G of zero content; and further let $f(x)$ be integrable in any sub-interval of (a, b) which contains no point of G either in its interior or at an end.

Let the set G be included in a definite number n of sub-intervals

$$\delta_1, \delta_2, \dots \delta_n,$$

each interval δ containing at least one point of G in its interior, so that the remaining part of (a, b) consists of a number of sub-intervals

$$\eta_1, \eta_2, \dots \eta_{\bar n},$$

which are free in their interiors and at their ends from points of G. Denote by $S_{\bar n}$ the sum of the integrals of $f(x)$ taken through all the sub-intervals η as these intervals are diminished and their number increased.

The number n can be so chosen that $\sum_1^n \delta$ is arbitrarily small; and therefore as n increases indefinitely $\sum_1^n \delta$ approaches the limit zero. Let a series of values of n be so chosen that $\Sigma\delta$ has a sequence of diminishing values $\epsilon_1, \epsilon_2, \dots$ which converge to the limit zero, and let $\bar n_1, \bar n_2, \bar n_3 \dots$ be the corresponding values of $\bar n$.

If the numbers $S_{\bar n_1}, S_{\bar n_2}, S_{\bar n_3} \dots$ form a convergent sequence of which S is*

* See Harnack, *Math. Annalen*, vol. XXIV, p. 220, where this definition is given in substance. See also Jordan, *Cours d'Analyse*, vol. II, p. 50, where a similar definition is given, except that the condition, that the set of points of infinite discontinuity should have zero content, is omitted.

the limit, and if S be independent of the particular choice of the intervals δ, then the number S is defined to be the improper integral of $f(x)$ in (a, b), and is denoted by $\int_a^b f(x)\,dx$.

The condition for the existence of the improper integral thus defined is that, corresponding to any arbitrarily small number ϵ, it must be possible to find a number ζ such that, if $\delta_1, \delta_2, \ldots \delta_n$ and $\delta_1', \delta_2', \ldots \delta'_{n'}$ be any two sets of intervals whatever of the type defined above, and such that

$$\sum_1^n \delta < \zeta, \quad \sum_1^{n'} \delta' < \zeta,$$

then the absolute difference of the corresponding sums $S_{\bar{n}}, S'_{\bar{n}'}$ is less than ϵ.

In case the improper integral $\int_a^b f(x)\,dx$ exist, it is the limit of the sum of the improper integrals of $f(x)$ through the set of sub-intervals complementary to the set of points G.

It is easily seen that the general definition of the improper integral is consistent with the definition which has been given for the case in which the number of points of infinite discontinuity is finite.

272. A definition of the improper integral has been given by de la Vallée-Poussin * which depends on a principle different from that employed in Harnack's definition.

Let $M_1, M_2, \ldots M_s \ldots$ and $N_1, N_2, \ldots N_s \ldots$ be two independent sequences of positive numbers each of which consists of continually increasing numbers which have no upper limit. Let a sequence of new functions be defined as follows:— If $f(x)$ be the given function, which has points of infinite discontinuity in the interval (a, b) for which it is defined, let $f_s(x)$ be defined so that

$$f_s(x) = f(x)$$

for all values of x which are such that

$$M_s \geqq f(x) \geqq -N_s; \text{ but } f_s(x) = M_s$$

for all values of x for which $f(x) > M_s$;

and $f_s(x) = -N_s$

for all values of x such that $f(x) < -N_s$;

if $f(x)$ be such that the integrals

$$\int_a^b f_1(x)\,dx, \quad \int_a^b f_2(x)\,dx, \quad \ldots \quad \int_a^b f_s(x)\,dx \ldots$$

all exist, and are such that they form a sequence which converges to a definite

* *Liouville's Journal*, ser. 4, vol. VIII, p. 427.

limit independent of the particular sequences $\{M\}$, $\{N\}$, *then that limit is defined to be the improper integral of* $f(x)$, *and is denoted by*

$$\int_a^b f(x)\,dx.$$

It will be observed that, whereas in Harnack's definition the improper integral is defined as the limit of a sequence of integrals of the same function taken through different domains, in de la Vallée-Poussin's definition the improper integral is defined as the limit of a sequence of integrals of different functions all taken through the same domain. It will however be seen later on, that this distinction is an unessential one.

ABSOLUTELY AND CONDITIONALLY CONVERGENT INTEGRALS.

273. *An improper integral* $\int_a^b f(x)$ *is said to be absolutely convergent if the improper integral* $\int_a^b |f(x)|\,dx$ *also exist; otherwise it is said to be conditionally convergent.*

It has been seen in § 256, from the definition of a proper integral, that all such integrals are absolutely convergent, and therefore the distinction between the two classes of integrals has reference to improper integrals only.

Every function $f(x)$ can be exhibited as the difference of two functions $f^+(x)$ and $f^-(x)$, defined so that $f^+(x)=f(x)$ when $f(x)>0$, $f^+(x)=0$ when $f(x)\leqq 0$; and $f^-(x)=-f(x)$ when $f(x)<0$, $f^-(x)=0$ when $f(x)\geqq 0$.

We have then $f(x)=f^+(x)-f^-(x)$, and $|f(x)|=f^+(x)+f^-(x)$. In an absolutely convergent improper integral both the functions $f^+(x)$, $f^-(x)$ possess improper integrals, but not so in the case of a conditionally convergent improper integral. For, the points of infinite discontinuity of the two functions $f(x)$, $|f(x)|$ being the same, the improper integrals of these two functions are the limits of proper integrals taken over sets of intervals which are the same for the two functions. It follows that $|f(x)|+f(x)$, $|f(x)|-f(x)$ have improper integrals, provided $|f(x)|$, $f(x)$ both have improper integrals.

The improper integrals defined in the manner of de la Vallée-Poussin are all absolutely convergent. For, since $\int_a^b f_s(x)\,dx$ has a limit as the two numbers M_s, N_s are independently increased indefinitely, and the integral is the sum of two parts, one dependent on M_s, and the other on N_s, it follows that each of these parts has separately a limit. Therefore the improper integral $\int_a^b f^+(x)\,dx$ exists, and similarly also $\int_a^b f^-(x)\,dx$; hence $\int_a^b |f(x)|\,dx$ exists in any case for which $\int_a^b f(x)$ is defined as an improper integral.

The definition of improper integrals in accordance with the method of Harnack applies both to absolutely, and to conditionally, convergent integrals, and is thus wider than the definition of de la Vallée-Poussin.

It has been pointed out by Schönflies that the condition that the set of points of infinite discontinuity must be of zero content, is deducible, in the case of de la Vallée-Poussin's definition, from the condition, contained in the definition, for the existence of the integral.

Since the existence of

$$\int_a^b |f(x)|\,dx$$

follows from that of

$$\int_a^b f(x)\,dx,$$

there will be no loss of generality if we suppose $f(x)$ to be everywhere positive. Now the condition for the existence of the integral as a finite number is that

$$\int_a^b \{f_{s+1}(x) - f_s(x)\}\,dx$$

should, as s is increased indefinitely, have the limit zero. Considering any convergent set of sub-divisions of (a, b), let $\delta_1, \delta_2, \ldots \delta_n$ be the sets of sub-intervals at any stage, and let σ be the sum of those δ's which contain points of infinite discontinuity of $f(x)$. In all these latter sub-intervals values of $f_s(x)$ equal to M_s, and values of $f_{s+1}(x)$ equal to M_{s+1}, occur; thus, in the sum whose limit defines the integral, the upper limit of $f_{s+1}(x) - f_s(x)$ is $M_{s+1} - M_s$, and the sum of the products of the intervals into the upper limits of the function in those intervals is $\geqq (M_{s+1} - M_s)\,\sigma$: hence the integral cannot have the limit zero unless σ converges to zero as the number n of intervals is increased indefinitely. Therefore the points of infinite discontinuity must form an unextended set, if the integral is to exist as a finite number.

274. In the case of absolutely convergent integrals it can be shewn that the definition of Harnack and that of de la Vallée-Poussin are in complete agreement. Since, in accordance with either definition, the existence of the absolutely convergent integral of $f(x)$ involves that of the improper integrals of $f^+(x)$, and $f^-(x)$, it is clearly sufficient to consider the case in which $f(x)$ is positive or zero at every point of the interval (a, b). Let us then assume that the function $f(x)$, which is never negative, has an improper integral in accordance with Harnack's definition. The set of points of infinite discontinuity of G is enclosed in a finite set of sub-intervals $\{\delta\}$, and the remaining part of (a, b) consists of a set of sub-intervals $\{\eta\}$. The sum $\Sigma\delta$ can be chosen so small that the integral of $f(x)$ through the intervals $\{\eta\}$ is less than Harnack's improper integral by less than an arbitrarily chosen positive number ζ. Let N be a positive number not less than the upper limit of $f(x)$ in all the intervals $\{\eta\}$, and let $f_n(x)$ be the function, corresponding to N, employed in de la Vallée-Poussin's definition. Let another set of sub-intervals $\{\delta'\}$ all interior to intervals of $\{\delta\}$, enclose all the points of infinite discontinuity of $f(x)$, and let $\{\eta'\}$ be the intervals complementary to these. The integral of $f(x)$ through $\{\eta'\}$ lies between the value of the integral through $\{\eta\}$, and that of Harnack's improper integral, and therefore differs from the latter by less than ζ.

It follows that the integral of $f(x)$ through the intervals obtained by removing the set $\{\delta'\}$ from the set $\{\delta\}$ is also $< \zeta$; and since $f_n(x) \leqq f(x)$, we see that the integral of $f_n(x)$ through the same set of intervals is $< \zeta$. From this we deduce that $\int f_n(x)\,dx$ taken through the intervals $\{\delta\}$ is less than $\zeta + N\Sigma\delta'$; and since this holds for an arbitrarily small value of $\Sigma\delta'$, N being fixed, we see that $\int f_n(x)\,dx$, taken through the intervals $\{\delta\}$, is $\leqq \zeta$. It now follows that $\int_a^b f_n(x)\,dx - \int_{\{\eta\}} f(x)\,dx \leqq \zeta$; and since ζ is arbitrarily small, n being sufficiently increased, it follows that $\int_a^b f_n(x)\,dx$ has a definite limit when n is indefinitely increased, and that this limit is Harnack's improper integral $\int_a^b f(x)\,dx$. It has thus been shewn that a function which has an absolutely convergent improper integral in accordance with Harnack's definition, has one also in accordance with the definition of de la Vallée-Poussin, the integrals having the same value in the two cases.

To prove the converse, we assume that

$$\int_a^b f_n(x)\,dx \quad \text{or} \quad \int_{\{\delta\}} f_n(x)\,dx + \int_{\{\eta\}} f(x)\,dx$$

has a definite limit as n is indefinitely increased and $\Sigma\delta$ indefinitely diminished. Since both the integrals are positive, it follows that $\int_{\{\eta\}} f(x)\,dx$, which increases as $\Sigma\delta$ is diminished, is less than a fixed finite number, and therefore has a definite upper limit. Therefore Harnack's improper integral exists, and it has been shewn above that it must then have the same value as de la Vallée-Poussin's. The two definitions have thus been shewn to be completely equivalent to one another, so far as they both apply to absolutely convergent integrals. The definition of Harnack is the wider, in that it applies to the case of non-absolutely convergent integrals.

EXISTENCE AND PROPERTIES OF ABSOLUTELY CONVERGENT IMPROPER INTEGRALS.

275. A definition of the improper integral of a function with infinite discontinuities having been given, it is necessary to investigate whether the limit employed in the definition really exists; and it is further necessary to discuss whether, in the case of the existence of the improper integral in (a, b), that improper integral shares the fundamental property of integrals that it exists for any and every sub-interval whatever of (a, b), and also exists for every set of such sub-intervals, and is in particular such as to satisfy the relation

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

It will be proved that the limit really exists in the case of absolutely convergent integrals, and that the improper integral then possesses the fundamental properties just specified. In the case in which the convergence is conditional, it appears that the improper integral when it exists possesses some but not all of the fundamental properties; in view of this fact, doubt has been thrown by some writers upon the appropriateness of regarding improper integrals with conditional convergence as really entitled* to the name of integral.

Consider first the case of a function $f(x)$ which has no negative values; so that in (a, b), $f(x) \geqq 0$. In this case it can be shewn that the numbers

$$S_{\bar{n}_1},\ S_{\bar{n}_2},\ S_{\bar{n}_3}\ \ldots$$

either increase beyond all limit, or have a limit S which is independent of the mode of formation of the successive sub-intervals.

Take sets of intervals

$$\begin{array}{l} \delta_1^{(1)},\ \delta_2^{(1)},\ \ldots\ \delta_{n_1}^{(1)}, \\ \delta_1^{(2)},\ \delta_2^{(2)},\ \ldots\ \delta_{n_2}^{(2)}, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \delta_1^{(m)},\ \delta_2^{(m)},\ \ldots\ \delta_{n_m}^{(m)}, \\ \ldots\ldots\ldots\ldots\ldots\ldots \end{array}$$

each set of which conaints all the points of infinite discontinuity; and let the corresponding sets of intervals which are free in their interiors and at their ends from the points of infinite discontinuity be

$$\begin{array}{l} \eta_1^{(1)},\ \eta_2^{(1)},\ \ldots\ \eta_{\bar{n}_1}^{(1)}, \\ \eta_1^{(2)},\ \eta_2^{(2)},\ \ldots\ \eta_{\bar{n}_2}^{(2)}, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \eta_1^{(m)},\ \eta_2^{(m)},\ \ldots\ \eta_{\bar{n}_m}^{(m)}, \\ \ldots\ldots\ldots\ldots\ldots\ldots \end{array}$$

Moreover, let the system of sets be so chosen that all points contained in the $\eta^{(m)}$ are also contained in the $\eta^{(m+1)}$. Since $f(x)$ is never negative, the sums $S_{\bar{n}_1}, S_{\bar{n}_2}, \ldots S_{\bar{n}_m} \ldots$ form a constantly increasing sequence; and thus the

* See Stolz, *Sitzungsberichte der kais. Akad. Wien*, vol. CVII, II a (1898), p. 207, and vol. CVIII, II a, p. 1234, also vol. CVII, p. 211. See also Stolz's work *Grundzüge der Diff. u. Integralrechnung*, part III, p. 278. In these writings there is a systematic treatment of the absolutely convergent improper integrals in accordance with Harnack's definition.

sequence must have a definite upper limit S, unless its terms increase indefinitely, in which case the improper integral is certainly not convergent. When S exists, we have to shew that it is independent of the particular sets of intervals used in obtaining it. The number m may be chosen so large that $S - S_{\bar{n}_m} < \epsilon$, where ϵ is a fixed arbitrarily chosen number.

Next, take any other set of successive sets of intervals which enclose the points of infinite discontinuity, leaving corresponding free intervals, and let $S_{m'}$ be the sum, at any stage, of the integrals taken through these free intervals: then compare $S_{\bar{n}_m}$ with $S_{m'}$. The two sums of integrals contain a number of integrals in common, namely integrals taken over those pieces of the interval (a, b) which are common to the sets of intervals belonging to $S_{\bar{n}_m}$ and $S_{m'}$; $S_{\bar{n}_m}$ may contain parts that do not belong to $S_{m'}$, these all forming parts of the intervals $\eta^{(m)}$; also $S_{m'}$ may contain parts that do not belong to $S_{\bar{n}_m}$. Now the difference $S_{m'} - S_{\bar{n}_m}$ is less than the sum of the integrals taken over these latter parts, all of which lie within the intervals $\delta^{(m)}$, and in all these parts the function $f(x)$ is limited; it follows that

$$S_{m'} - S_{\bar{n}_m} < P\Sigma\delta^{(m)},$$

where P is a number which does not increase as m is increased. Now m may be chosen so great that

$$\Sigma\delta^{(m)} < \frac{\epsilon}{P},$$

and then $$S_{m'} - S_{\bar{n}_m} < \epsilon,$$

hence $$S_{m'} < S + \epsilon;$$

and since ϵ is arbitrarily small, we have

$$S_{m'} \leqq S;$$

thus $S_{m'}$ cannot be greater than S. We have again, by similar reasoning

$$S_{\bar{n}_m} - S_{m'} < Q\Sigma\delta',$$

where Q is a number which does not increase as the number m' is increased, depending as it does on the upper limits of $f(x)$ for intervals all of which lie in the $\eta^{(m)}$; hence when m' is sufficiently great

$$S_{\bar{n}_m} - S_{m'} < \epsilon, \quad \text{or} \quad S_{m'} > S_{\bar{n}_m} - \epsilon > S - \epsilon.$$

It thus appears that the second system of divisions can be so far advanced that $S_{m'}$ lies between $S - \epsilon$ and $S + \epsilon$, where ϵ is arbitrarily small; and hence $S_{m'}$ has S for its limit. The existence of the improper integral is now established for a function $f(x) \geqq 0$, provided there be no divergence.

The theorem proved may be immediately extended to shew that the integral $\int_a^b f(x)\,dx$ exists, provided

$$\int_a^b |f(x)|\,dx$$

exist, that is provided the convergence be absolute.

Replace $f(x)$ by the difference of the two functions $f^+(x)$ and $f^-(x)$; then both the integrals

$$\int_a^b f^+(x)\,dx, \quad \int_a^b f^-(x)\,dx$$

exist, unless the sums of integrals of which they are the limits increase indefinitely.

Now $$\int_a^b |f(x)|\,dx = \int_a^b f^+(x)\,dx + \int_a^b f^-(x)\,dx;$$

hence if $|f(x)|$ have an integral, both the functions $f^+(x)$, $f^-(x)$ have integrals, and thus $f(x)$ has also an integral. The existence of absolutely convergent improper integrals has thus been established.

276. *If c be a number such that $a < c < b$, and if $f(x)$ have an absolutely convergent improper integral in (a, b), then it has also such integrals in (a, c) and in (c, b), and the sum of the two latter integrals is equal to the former one.*

Consider a system of intervals $\{\delta\}$, $\{\eta\}$ as in § 275. If the point c lies in an interval δ it does not affect the sums of the proper integrals through the intervals η; but if, at any stage of the limiting process, the point c comes to lie within an interval η, it divides it into two parts. Clearly, the sum of the integrals of $|f(x)|$ taken through those intervals η which are on the left of c, and through that part of the interval containing c which lies on the left of c, is less than the sum of the integrals of $|f(x)|$ through all the intervals η. The same is true of the sum of the integrals of $|f(x)|$ through all those intervals η which lie on the right of c, and through that part of the interval containing c which lies to the right of c.

Thus, since the integral of $|f(x)|$ through all the intervals η lies below a fixed limit, the same is true of the two parts into which the integral is divided by the point c; and thus the two integrals

$$\int_a^c |f(x)|\,dx, \quad \int_c^b |f(x)|\,dx$$

exist; and therefore also

$$\int_a^c f(x)\,dx, \quad \int_c^b f(x)\,dx$$

exist.

The splitting up of the sum of the integrals of $f(x)$ through the intervals η into two parts does not affect that sum; hence also in the limit we have

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

A corollary from this theorem is that *if $f(x)$ have an absolutely convergent improper integral in (a, b), it is also integrable in any interval (a', b') which forms part of (a, b).*

277. If $\Delta_1, \Delta_2, \dots \Delta_n, \dots$ be a sequence of non-overlapping intervals contained in (a, b), in descending order of length, the sum of the integrals of $f(x)$ taken through $\Delta_1, \Delta_2, \dots \Delta_n$ converges to a definite limit as n is increased indefinitely, provided

$$\int_a^b f(x)\,dx$$

converges absolutely.

Consider the two functions $f^+(x), f^-(x)$ defined as in § 273. The function $f^+(x)$ is integrable in each of the intervals Δ, and the sum of the integrals of $f^+(x)$ through the intervals $\Delta_1, \Delta_2, \dots \Delta_n$ is positive and does not diminish as n is increased; also this sum never exceeds the integral of $f^+(x)$ through (a, b). It follows that the sum of the integrals of $f^+(x)$ through

$$\Delta_1, \Delta_2, \dots \Delta_n,$$

converges to a definite limit as n is increased indefinitely. The same is true of $f^-(x)$; and hence $f(x)$, which is $f^+(x) - f^-(x)$, is such that the sum of its integrals converges to a definite limit when the number of the intervals Δ is increased indefinitely.

It has thus been shewn that, *if $f(x)$ have an absolutely convergent improper integral in (a, b), it has also an improper integral through any portion of (a, b) which consists of a finite, or of an infinite, number of continuous intervals.*

278. *If H be a set of points in (a, b) of zero content, so that the points of H can be enclosed in intervals $\theta_1, \theta_2, \dots \theta_p$ whose sum is arbitrarily small, then the integral of $f(x)$ taken through the intervals $\theta_1, \theta_2, \dots \theta_p$ has the limit zero, as the sum of the intervals converges to zero, their number being indefinitely increased, $f(x)$ having an absolutely convergent improper integral in (a, b).*

The points of infinite discontinuity of $f(x)$ can be enclosed in a set of intervals $\delta_1, \delta_2, \dots \delta_n$ such that the integral of $|f(x)|$ taken through these intervals is $< \epsilon$. Let $\phi_1, \phi_2, \dots \phi_r$ be the parts of the intervals $\theta_1, \theta_2, \dots \theta_p$ which are common with the intervals $\eta_1, \eta_2, \dots \eta_{n_m}$ which remain in (a, b) when the δ's are removed. The absolute value of the integral of $f(x)$ taken through the intervals $\theta_1, \theta_2, \dots \theta_p$ is less than $\epsilon + U\sum_1^r \phi$, or than $\epsilon + U\sum_1^p \theta$, where U is the finite upper limit of $|f(x)|$ in all the intervals η. Having fixed the intervals $\delta_1, \delta_2, \dots \delta_n$, we can choose the intervals θ so that $\sum_1^p \theta < \epsilon/U$; thus the absolute value of $\int f(x)\,dx$ taken through the intervals θ is then $< 2\epsilon$, which is arbitrarily small. The theorem is therefore established.

It follows that, in the definition of the improper integral $\int_a^b f(x)\,dx$ as the limit of the sum of the proper integrals through the intervals η, we may suppose the neighbourhoods of the points of H to be removed from (a, b), these neighbourhoods being so chosen that their sum converges to zero as the sum of the proper integrals converges to its limit, the improper integral of $f(x)$ in (a, b).

The theorem may also be stated in the form that, *if H be any set of points of zero content, then in applying Harnack's definition, the set H may be added to the set G of points of infinite discontinuity, without altering the value of the integral.*

From this theorem we may deduce that *if $f(x)$, $\psi(x)$ have both absolutely convergent improper integrals in (a, b), their sum $f(x)+\psi(x)$ has an absolutely convergent improper integral in (a, b).*

In relation to $\int_a^b f(x)\,dx$, the points of infinite discontinuity of $\psi(x)$ form a set H such as is contemplated in the foregoing theorem; thus in the definition of $\int_a^b f(x)\,dx$ we may exclude the neighbourhoods of the points H. A similar remark applies to $\int_a^b \psi(x)\,dx$. The points of infinite discontinuity of $f(x)+\psi(x)$ consist in general of the sets for $f(x)$ and for $\psi(x)$ together. It therefore follows that $\int_a^b\{f(x)+\psi(x)\}\,dx$ exists, and is identical with

$$\int_a^b f(x)\,dx+\int_a^b \psi(x)\,dx.$$

The following theorem has been incidentally established:—

If 2ϵ be an arbitrarily chosen positive number, then a positive number η, dependent on ϵ, can be determined, such that, for any finite set of non-overlapping intervals whatever, whose sum is less than η, the absolute value of the sum of the integrals of $f(x)$ through the intervals is less than 2ϵ; it being assumed that $f(x)$ has an absolutely convergent integral in (a, b).

279. *If $f(x)$, $\psi(x)$ have both absolutely convergent improper integrals in (a, b), and if the two functions have no points of infinite discontinuity in common, then the product $f(x)\psi(x)$ has an absolutely convergent improper integral in (a, b).*

The infinities of $f(x)$ may be included in intervals $\delta_1, \delta_2, \ldots \delta_n$, and those of $\psi(x)$ in intervals $\delta_1', \delta_2', \ldots \delta_m'$, such that no interval δ encroaches on any interval δ'. Let U be the upper limit of $f(x)$ in all the intervals δ', and let U' be the upper limit of $\psi(x)$ in all the intervals δ; thus U, U' are definite numbers. In the intervals δ, $|f(x)\psi(x)|$ never exceeds $U'\,|f(x)|$, and hence since $|f(x)|$ is integrable in the intervals δ, having an improper integral in each of these intervals, in accordance with the definition in § 271, it is clear that $|f(x)\psi(x)|$ is also integrable in these intervals δ; and $\int|f(x)\psi(x)|\,dx$

taken through the intervals δ is $\leq U' \int |f(x)|\,dx$ taken through the same intervals. Similarly it may be shewn that $|f(x)\psi(x)|$ is integrable in the intervals δ'. It follows that $|f(x)\psi(x)|$ is, under the conditions in the enunciation, integrable in the whole interval (a, b); and that therefore $f(x)\psi(x)$ has an absolutely convergent improper integral in that interval.

280. *If $f(x)$ have an absolutely convergent improper integral in (a, b), then the improper integral $\int_a^x f(x)\,dx$ is a continuous function of the upper limit x.*

This theorem is the extension of that of § 258 to the case of absolutely convergent improper integrals.

Let $F(x)$ denote $\int_a^x f(x)\,dx$, which has been shewn, in § 276, to exist for $a \leq x \leq b$. We have $F(x+h) - F(x) = \int_x^{x+h} f(x)\,dx$; if then x is not a point of infinite discontinuity of $f(x)$, h may be so chosen that $(x, x+h)$ does not contain any such points, and in that case $|F(x+h) - F(x)| \leq |h|\,U$, where U denotes the upper limit of $|f(x)|$ in the interval $(x, x+h)$; since U does not increase as h is diminished, it follows that, corresponding to a fixed number ϵ, a value of h, say h_1, can be found such that $|F(x+h) - F(x)| < \epsilon$, for $|h| \leq |h_1|$; hence $F(x)$ is continuous at x.

In case x be a point of infinite discontinuity of $f(x)$, we may enclose all such points in a finite set of intervals such that the integral of $|f(x)|$ taken through all of them is $< \epsilon$. Since $\left|\int_x^{x+h} f(x)\,dx\right| \leq \int_x^{x+h} |f(x)|\,dx$, if we choose h so small that $(x, x+h)$ is entirely within that interval of the set which contains the point x, we see that $|F(x+h) - F(x)| < \epsilon$; and hence the point x is a point of continuity of $F(x)$.

NON-ABSOLUTELY CONVERGENT IMPROPER INTEGRALS.

281. It has been shewn that an absolutely convergent improper integral, when it exists, possesses the fundamental properties which belong to a proper integral, viz. that the function is also integrable in any part of the interval which is either continuous or which consists of a finite or infinite number of continuous portions, and that $\int_a^b f(x)\,dx = \int_a^x f(x)\,dx + \int_x^b f(x)\,dx$. Further it has been shewn that $\int_a^x f(x)\,dx$ is a continuous function of the upper limit x.

If we apply the definition of § 271, to the case in which $\int_a^b |f(x)|\,dx$ does not exist, it has not been shewn that the sum of the integrals of $f(x)$ through the intervals η which remain when the points of infinite discontinuity are enclosed in a set of intervals δ, necessarily either converges to a definite limit, or increases indefinitely. Further, if the limit which defines $\int_a^b f(x)\,dx$ in any particular case actually exists, it has not been shewn that $f(x)$ is necessarily integrable in (a, x), or in general in every interval contained in (a, b), the proof in § 276, depending essentially upon the assumption that $|f(x)|$ is integrable in (a, b). It is thus a matter for further investigation whether a non-absolutely convergent improper integral defined in Harnack's manner necessarily possesses the fundamental properties which would justify us in regarding it as an extension of the conception of a proper integral. The definition of de la Vallée-Poussin, in § 272, is applicable only to absolutely convergent improper integrals.

Under these circumstances the definition of a non-absolutely convergent improper integral has, by some writers*, been restricted to the case in which the set G of points of infinite discontinuity of the function is enumerable and of the first species. The mode of definition usually applied in this case will first be briefly considered, before the more general definition of Harnack is considered.

First let us suppose that G consists of a finite number of points

$$c_1, c_2, \ldots c_n,$$

then as in § 270,

$$\int_{c_{r-1}}^{c_r} f(x)\,dx$$

is defined to be the limit of

$$\int_{c_{r-1}+\epsilon}^{c_r-\epsilon'} f(x)\,dx,$$

as ϵ, ϵ' independently of one another converge to zero, on the assumption that this limit exists.

If the integrals

$$\int_a^{c_1} f(x)\,dx, \quad \int_{c_1}^{c_2} f(x)\,dx, \ldots \quad \int_{c_n}^{b} f(x)\,dx,$$

all exist in accordance with this definition, their sum is defined to be the improper integral

$$\int_a^b f(x)\,dx.$$

* See Du Bois Reymond, *Crelle's Journal*, vol. LXXIX, pp. 36 and 45, also Dini, *Grundlagen*, p. 404.

This definition is applicable, whether $|f(x)|$ be integrable in accordance with it, or not; and it thus defines non-absolutely convergent integrals in the case considered.

Next, let us suppose that G is of the first species and of the first order. In this case G' consists of a finite number of points $e_1, e_2, \dots e_r$.

If all the improper integrals

$$\int_a^{e_1-\epsilon_1} f(x)\,dx, \quad \int_{e_1+\epsilon_1'}^{e_2-\epsilon_2} f(x)\,dx, \quad \int_{e_2+\epsilon_2'}^{e_3-\epsilon_3} f(x)\,dx, \dots,$$

each of which falls under the last case, have each a definite limit as $\epsilon_1, \epsilon_1', \epsilon_2, \epsilon_2', \dots$ converge, independently of one another, to the limit zero, then the sum of these limits is taken to define the integral

$$\int_a^b f(x)\,dx.$$

It is clear that this definition admits of extension to the case in which G is of the first species and of any order. It is also clear that an integral in (a, b), which exists in accordance with this definition, entails the existence of the integral in (a, x), and in any continuous interval contained in (a, b); and further the truth of the theorem

$$\int_a^b f(x)\,dx = \int_a^x f(x)\,dx + \int_x^b f(x)\,dx$$

is assured.

The definition has been extended by Schönflies* to the case in which G is enumerable but possesses derivatives of transfinite order.

In the case in which $f(x)$ is absolutely integrable in accordance with Harnack's definition, and in which G is of the first species, it can be easily shewn that Harnack's definition reduces to the one here given.

It should be observed that, in the case of a non-absolutely convergent improper integral which has an infinite set of points of infinite discontinuity, the theorem that the function is integrable through any set of intervals contained in the interval of integration, does not in general hold; so that such improper integrals are not in this respect on a parity with proper integrals. For it may be possible to choose an infinite set of intervals so that $f(x)$ is everywhere positive in them; and then the sum of the integrals of $f(x)$ through these intervals does not in general converge to a finite value, the existence of the integral in (a, b) depending essentially on the cancelling of the integrals through those parts of (a, b) in which $f(x)$ is positive, with the integrals through those parts in which $f(x)$ is negative. The two integrals

$$\int_a^b f^+(x)\,dx, \quad \int_a^b f^-(x)\,dx$$

* See his *Bericht*, p. 185; a similar definition has also been employed by de la Vallée-Poussin, *loc. cit.*, p. 453.

have no finite values, although

$$\int_a^b \{f^+(x) - f^-(x)\}\, dx$$

may have a definite finite value.

That $F(x) = \int_a^x f(x)\, dx$ is a continuous function of the upper limit x, in the case when the integral is a non-absolutely convergent improper integral, in accordance with the definition here given, can be shewn as follows:—

Since $$F(x+h) - F(x) = \int_x^{x+h} f(x)\, dx;$$

if x be a point of infinite discontinuity of $f(x)$, we know that the integral on the right-hand side has the limit zero, when h is indefinitely diminished either through positive or through negative values; and hence a value h_1 can be found such that $|F(x+h) - F(x)| < \epsilon$, for $|h| < h_1$.

In case x be not a point of infinite discontinuity of $f(x)$, the proof is identical with that which has been given for the case of a proper integral.

EXAMPLES.

1. Let $f(x)$ denote a function which is integrable in every interval (a, b), where $0 < a < b$; and let $f(x)$, be in the neighbourhood of the point 0, of the form $\frac{\phi(x)}{x^k}$, where k is positive, and $\phi(x)$ is a limited function of constant sign.

We have

$$\left|\int_\epsilon^{\epsilon'} \frac{\phi(x)}{x^k}\, dx\right| < A \int_\epsilon^{\epsilon'} \frac{dx}{x^k} < \frac{A}{1-k}[\epsilon'^{1-k} - \epsilon^{1-k}],$$

where A is some positive number.

If $0 < k < 1$, it is clear that $\int_\epsilon^{\epsilon'} \frac{\phi(x)}{x^k}\, dx$ is arbitrarily small for a sufficiently small value of $\epsilon' (> \epsilon)$, and therefore the improper integral $\int_0^c f(x)\, dx$ exists, being convergent at the point $x=0$. If $k \geqq 0$, the improper integral does not exist.

2. Let $f(x)$ be, in the neighbourhood on the right of the point 0, of the form $\frac{\phi(x)}{x[\log x]^{1+p}}$, where p is positive, and $\phi(x)$ satisfies the same condition as in Ex. 1.

We have

$$\left|\int_\epsilon^{\epsilon'} \frac{\phi(x)}{x[\log x]^{1+p}}\, dx\right| < \frac{A}{p}\{[\log \epsilon]^{-p} - [\log \epsilon']^{-p}\};$$

and thus the improper integral $\int_0^c f(x)\, dx$ exists, being absolutely convergent.

3. $\int \frac{\tan x}{x}\, dx$, taken through any interval which contains a point of infinite discontinuity of $\tan x$, does not exist.

For $$\int_{\frac{\pi}{2}-\epsilon}^{\frac{\pi}{2}-\epsilon'} \frac{\tan x}{x}\,dx > \frac{2}{\pi}\log\frac{\sin\epsilon}{\sin\epsilon'},$$

and this is arbitrarily great for a sufficiently great value of ϵ/ϵ'; thus the integral does not converge at the point $x=\frac{1}{2}\pi$. The integral possesses however a principal value at the point $\frac{1}{2}\pi$. For the sum of the integrals taken through the intervals $(\frac{1}{2}\pi-\epsilon,\ \frac{1}{2}\pi-\epsilon')$ and $(\frac{1}{2}\pi+\epsilon',\ \frac{1}{2}\pi+\epsilon)$ is

$$\int_{\epsilon'}^{\epsilon}\cot x\,.\,\frac{2x}{\frac{1}{4}\pi^2-x^2}\,dx < 2\epsilon\,(\tfrac{1}{4}\pi^2-\epsilon^2)^{-1}(1-\tfrac{1}{6}\epsilon^2)^{-1},$$

and this converges to 0, with ϵ.

4. The function $\cos(e^{\frac{1}{x}})+\frac{1}{x}e^{\frac{1}{x}}\sin(e^{\frac{1}{x}})$ oscillates between indefinitely great positive and negative values, in the neighbourhood of the point $x=0$. For every value of x except $x=0$, the function $=\frac{d}{dx}\{x\cos(e^{\frac{1}{x}})\}$.

Also $$\int_{\epsilon'}^{\epsilon}\frac{d}{dx}\{x\cos(e^{\frac{1}{x}})\}\,dx=\epsilon\cos(e^{\frac{1}{\epsilon}})-\epsilon'\cos(e^{\frac{1}{\epsilon'}}),$$

where $\epsilon>\epsilon'>0$. It thus appears that the integral of the function converges at the point $x=0$; and therefore the function is integrable in an interval containing that point.

282. The general theory of improper definite integrals, both those which converge absolutely, and those which converge non-absolutely, defined according to Harnack's definition, has been treated by E. H. Moore*, who has also considered other definitions of such improper integrals.

It has been shewn in § 276, that if $f(x)$ have an improper integral in (a, b), in accordance with Harnack's definition, and such integral be absolutely convergent, then $f(x)$ is also integrable in any interval (a', b') which is part of (a, b). It will now be shewn that this holds whether the improper integral converges absolutely or not†.

Let $\{\delta\}$ denote a finite set of intervals enclosing all the points of infinite discontinuity of $f(x)$, each interval of the set enclosing at least one such point; we may denote by $f_\delta(x)$ a function which is zero at all points interior to the intervals $\{\delta\}$, and is at every other point of (a, b) equal to $f(x)$. The corresponding function for any other such set of intervals $\{\delta'\}$ may be denoted by $f_{\delta'}(x)$.

The condition stated in § 271, for the existence of the improper integral $\int_a^b f(x)\,dx$, may be expressed in the form that, corresponding to an arbitrarily fixed positive number ϵ, it shall be possible to fix a number ζ, such that, for any two sets of intervals $\{\delta\}$, $\{\delta'\}$, such that $\Sigma\delta<\zeta$, $\Sigma\delta'<\zeta$, the condition

$$\left|\int_a^b f_\delta(x)\,dx-\int_a^b f_{\delta'}(x)\,dx\right|<\epsilon$$

may be satisfied.

* *Trans. Amer. Math. Soc.*, vol. II, 1901, p. 296, and a second paper, p. 459.

† This is contrary to a statement made by Stolz; see *Grundzüge*, vol. III, p. 277.

Assuming that this condition is satisfied for every value of ϵ, it will be shewn that, for every pair of points a', b' in (a, b), the condition

$$\left| \int_{a'}^{b'} f_\delta(x)\,dx - \int_{a'}^{b'} f_{\delta'}(x)\,dx \right| < \epsilon$$

is satisfied, provided $\{\delta\}$, $\{\delta'\}$ are any two sets of intervals of the prescribed kind, and such that $\Sigma\delta$, $\Sigma\delta'$ are each less than $\frac{1}{2}\zeta$.

Let it be assumed that, if possible, a', b', $\{\delta\}$, $\{\delta'\}$ can be so determined, subject to the conditions $\Sigma\delta < \frac{1}{2}\zeta$, $\Sigma\delta' < \frac{1}{2}\zeta$, that

$$\left| \int_{a'}^{b'} f_\delta(x)\,dx - \int_{a'}^{b'} f_{\delta'}(x)\,dx \right| \geqq \epsilon.$$

It will then be shewn that finite sets $\{\delta^{(2)}\}$, $\{\delta^{(3)}\}$ can be determined, each of total length less than ζ, and each containing the set of points of infinite discontinuity, for which

$$\left| \int_a^b f_{\delta^{(2)}}(x)\,dx - \int_a^b f_{\delta^{(3)}}(x)\,dx \right| \geqq \epsilon;$$

and since this is contrary to the hypothesis, the impossibility of the above assumption will have been demonstrated.

To define $\{\delta^{(2)}\}$, $\{\delta^{(3)}\}$, we take any interval of $\{\delta\}$ within (a', b'), as an interval of $\{\delta^{(2)}\}$; and any interval of $\{\delta'\}$ within (a', b'), as an interval of $\{\delta^{(3)}\}$. Further, we take for the parts of $\{\delta^{(2)}\}$ and $\{\delta^{(3)}\}$ within (a, a') and (b', b) the set of those intervals which are common to the parts of $\{\delta\}$ and $\{\delta'\}$ that lie in (a, a') and (b', b).

In case a' is contained in intervals (α, β), (α', β'), of $\{\delta\}$ and of $\{\delta'\}$ respectively, we take (a', β) and (a', β') as intervals of $\{\delta^{(2)}\}$ and of $\{\delta^{(3)}\}$ respectively, where $\alpha' > \alpha$. A similar specification will refer to b'.

It is now clear that, in accordance with these definitions of $\{\delta^{(2)}\}$ and $\{\delta^{(3)}\}$, we have $f_\delta(x) = f_{\delta^{(2)}}(x)$, and $f_{\delta'}(x) = f_{\delta^{(3)}}(x)$, if x is within (a', b'); and $f_{\delta^{(2)}}(x) = f_{\delta^{(3)}}(x)$, if x is within (a, a'), or within (b', b).

It follows that

$$\int_a^b f_{\delta^{(2)}}(x)\,dx - \int_a^b f_{\delta^{(3)}}(x)\,dx = \int_{a'}^{b'} f_\delta(x)\,dx - \int_{a'}^{b'} f_{\delta'}(x)\,dx,$$

and thence that

$$\left| \int_a^b f_{\delta^{(2)}}(x)\,dx - \int_a^b f_{\delta^{(3)}}(x)\,dx \right| \geqq \epsilon;$$

moreover it is clear from the mode of construction of $\{\delta^{(2)}\}$ and $\{\delta^{(3)}\}$, that $\Sigma\delta^{(2)} < \zeta$, $\Sigma\delta^{(3)} < \zeta$. The impossibility in question has therefore been demonstrated.

Since for every pair of numbers a', b' such that $a \leqq a' < b' \leqq b$, corresponding to any arbitrarily chosen number ϵ, a number $\frac{1}{2}\zeta$ can be found such that

$$\left|\int_{a'}^{b'} f_\delta(x)\,dx - \int_{a'}^{b'} f_{\delta'}(x)\,dx\right| < \epsilon,$$

for every pair of sets of intervals $\{\delta\}$, $\{\delta'\}$, enclosing the points of infinite discontinuity, at least one such point being contained in each interval of either set, and such that $\Sigma\delta < \frac{1}{2}\zeta$, $\Sigma\delta' < \frac{1}{2}\zeta$, it follows that $\int_{a'}^{b'} f(x)\,dx$ exists.

Moreover, since ζ is independent of a', b', we have established the following theorem:—

If $\int_a^b f(x)\,dx$ exist as an improper integral, in accordance with Harnack's definition, then $\int_{a'}^{b'} f(x)\,dx$ also exists, where a', b' are such that $a \leqq a' < b' \leqq b$; and the convergence of this integral is uniform for all values of a' and b'.

The last part of this theorem expresses the fact that

$$\left|\int_{a'}^{b'} f(x)\,dx - \int_{a'}^{b'} f_\delta(x)\,dx\right| \leqq \epsilon,$$

provided $\Sigma(\delta) < \frac{1}{2}\zeta$, for every value of a' and b'; the number ζ depending on the arbitrarily chosen number ϵ.

The theorem $\int_a^{b'} f(x)\,dx + \int_{b'}^{c'} f(x)\,dx = \int_{a'}^{c'} f(x)\,dx$ *is valid.*

This follows from the corresponding theorem for the proper integrals of $f_\delta(x)$; for it appears that the expressions on the two sides of the equation differ from one another by 2ϵ at most; and since ϵ is arbitrarily small, their equality is established.

Since the existence of the integral of $f(x)$ in any sub-interval (a', b') of (a, b) has been shewn to be a necessary consequence of the existence of the integral in (a, b), it is clear that the integral of $f(x)$ taken through any finite set of non-overlapping intervals contained in (a, b) also exists; being the sum of the integrals taken through the separate intervals. However, if a non-finite set of non-overlapping intervals be taken in (a, b), it is not in general true that the sum of the integrals of $f(x)$ through these intervals converges to a definite number, unless the integral of $f(x)$ is absolutely convergent, which case has been treated in § 277. It will in fact be shewn, by means of an example, that the property in question, that $f(x)$ is integrable through a non-finite set of intervals in (a, b), does not appertain to non-absolutely convergent integrals, and must be regarded as peculiar to absolutely convergent integrals. This does not however seem a sufficient reason for refraining from applying the term "integral" to non-absolutely convergent improper integrals.

283. The following theorem contains the necessary and sufficient conditions for the existence of the improper integral of a function $f(x)$ in an interval (a, b), in which the set G of points of infinite discontinuity of $f(x)$ exists.

The complementary intervals of G being denoted by (a_ν, b_ν), the necessary and sufficient conditions for the existence of $\int_a^b f(x)\,dx$ *are*

(1) *that all the integrals* $\int_{a_\nu}^{b_\nu} f(x)\,dx$ *shall exist, each such integral being defined as the limit of* $\int_{a_\nu+\epsilon}^{b_\nu-\epsilon'} f(x)\,dx$, *when* ϵ, ϵ' *converge independently to the limit zero, and*

(2) *that* $\omega_1 + \omega_2 + \ldots + \omega_\nu$ *shall converge to a definite number, as ν is indefinitely increased; where ω_ν denotes the fluctuation of* $\int_{a_\nu}^{x} f(x)\,dx$ *in the interval* (a_ν, b_ν).

Moreover, when the conditions (1) *and* (2) *are satisfied, the sum*

$$\sum_{\nu=1}^{\nu} \int_{a_\nu}^{b_\nu} f(x)\,dx$$

is convergent, and its limit, as ν is indefinitely decreased, is $\int_a^b f(x)\,dx$.

For the proof of this theorem, which is due to E. H. Moore, reference must be made to the original memoir*.

It will be observed that in Harnack's definition of an improper integral, the set of intervals $\{\delta\}$ which are of arbitrarily small sum, and which enclose the points of infinite discontinuity of the function, have been so chosen that each interval δ contains at least one of these points. If this latter condition were omitted from the definition, the amended definition would admit only of the existence of absolutely convergent improper integrals. An integral thus defined has been named by E. H. Moore a *broad* integral, in contradistinction to the *narrow* integral as given by Harnack's original definition. It is unnecessary here to shew that a broad integral is necessarily absolutely convergent, because the corresponding definition for double integrals will be fully considered below. The broad integrals are a special case of the narrow ones; those narrow integrals which are not broad ones are the non-absolutely convergent integrals.

284. A method will now be given of constructing a function $f(x)$ which is continuous at every point of the interval (a, b), except at the point b at

* *loc. cit.*, p. 324.

which the function has an infinite discontinuity of such a character that $\int_a^b f(x)\,dx$ converges non-absolutely.

Let a sequence of intervals $(a_1, b_1), (a_2, b_2), \dots (a_n, b_n) \dots$ be defined in the interval (a, b), such that no two of the intervals overlap, and that b is the limiting-point of each of the sequences $(a_1, a_2, \dots a_n \dots)$, $(b_1, b_2, \dots b_n, \dots)$. Let $u_1 + u_2 + \dots + u_n + \dots$ denote a non-absolutely convergent arithmetic series (see Chap. VI). In (a_n, b_n), let $f(x)$ be defined so as to be continuous in that interval, and everywhere of the same sign, and let $f(x)$ vanish at a_n and b_n. Further, let $f(x)$ be so chosen in the interval, that

$$\int_{a_n}^{b_n} f(x)\,dx = u_n.$$

At all points of (a, b) exterior to all the intervals (a_n, b_n), let $f(x) = 0$.

The function $f(x)$ so defined is continuous in (a, b), except at b.

In (a_n, b_n), the function $|f(x)|$ has a maximum greater than $|u_n|/(b_n - a_n)$, and therefore $f(x)$ has indefinitely great positive and negative values in every neighbourhood of the point b.

We have now

$$\int_a^x f(x)\,dx = \sum_{r=1}^{r=n} u_n + \theta u_{n+1},$$

if x lies in the interval (b_n, b_{n+1}); where θ is some proper fraction.

Now the improper integral $\int_a^b f(x)\,dx$ is defined by $\lim\limits_{x=b} \int_a^x f(x)\,dx$, and its value is therefore the limiting sum of the series $u_1 + u_2 + \dots + u_n + \dots$. It is further clear that $\int_a^b |f(x)|\,dx$ does not exist, since the series

$$|u_1| + |u_2| + \dots + |u_n| + \dots$$

is not convergent.

This case may be employed to illustrate the fact that the non-absolutely convergent improper integral is not necessarily the limit of the sum of the integrals taken through a set of intervals which in the limit converges to the whole interval of integration; and thus that such an integral is not a broad integral.

Let the integral of $f(x)$ be taken through the intervals

$$(a, b_m), (a_{p_1}, b_{p_1}), (a_{p_2}, b_{p_2}) \dots (a_{p_r}, b_{p_r}),$$

where $p_1, p_2, \dots p_r$ are increasing numbers all $> m$, such that $u_{p_1}, u_{p_2}, \dots u_{p_r}$ are all of the same sign. It is clear that m may be so chosen that $\int_a^{b_m} f(x)\,dx$

is arbitrarily near to $\int_a^b f(x)\,dx$; then, for such a fixed value of m, the numbers $p_1, p_2, \dots p_r$ may be so chosen that $u_{p_1} + u_{p_2} + \dots u_{p_r}$ is as large as we please, since the series Σu_n is non-absolutely convergent. As m is increased indefinitely, the set of intervals (a, b_m), $(a_{p_1}, b_{p_1}) \dots (a_{p_r}, b_{p_r})$ converges to the whole interval (a, b), the total length of the complementary part of (a, b) diminishing indefinitely, and yet the sum of the integrals of $f(x)$ taken through the set of intervals is divergent.

This example may be used to illustrate the fact that the theorem established in § 278, for absolutely convergent integrals, does not hold for non-absolutely convergent integrals. It is not in fact true that, in defining $\int_a^b f(x)\,dx$, the set of points $a_1, a_2, \dots a_n \dots b_1, b_2, \dots b_n, \dots b$, which is of zero content, may be excluded by enclosing these points in a set of intervals of arbitrarily small sum. For we may include all the points $a_1, a_2, \dots a_m$, $b_1, b_2, \dots b_m$ which occur in (a, b_m) in a finite set of intervals, so that when these are excluded from the domain of integration of $\int_a^{b_m} f(x)\,dx$, that integral is altered by an arbitrarily small amount.

Again we may shorten each of the intervals $(a_{p_1}, b_{p_1}) \dots (a_{p_r}, b_{p_r})$ at each end, so that the sum of the integrals taken through these intervals is diminished by an arbitrarily small amount. All the points $a_1, a_2, \dots$, $b_1, b_2, \dots b$ are now included in a finite set of intervals, such that the integral of $f(x)$ taken through the complementary intervals has an arbitrarily great sum. These complementary intervals consist of those intervals which have been obtained by shortening $(a_{p_1}, b_{p_1}) \dots (a_{p_r}, b_{p_r})$, and of the parts of (a, b_m) which remain when the points $a_1, b_1, a_2, b_2, \dots b_m$ have been included in a suitable set of intervals.

Let $\phi(x)$ be an improper integral for which all the points $a_1, a_2, \dots$, $b_1, b_2, \dots b$ are points of infinite discontinuity; and thus $\int_a^b \phi(x)\,dx$ may exist in accordance with Harnack's definition. Also $\int_a^b f(x)\,dx$ exists, as defined above, having its single point of infinite discontinuity at b. It appears however that $\int_a^b \{f(x) + \phi(x)\}\,dx$ does not exist, because $f(x) + \phi(x)$ has infinite discontinuities at all the points a_n, b_n, and at b, and its existence would imply that in defining $\int_a^b f(x)\,dx$ we could employ sets of intervals which exclude not only the point b, but also all the points a_n, b_n.

THE FUNDAMENTAL THEOREM OF THE INTEGRAL CALCULUS FOR THE CASE OF IMPROPER INTEGRALS.

285. The theorem of § 260, that if $f(x)$ be integrable in (a, b), and $F(x)$ be the corresponding integral function, any one of the four derivatives $DF(x)$ of $F(x)$ is integrable in (a, b), and has $F(x)$ for its integral function, is applicable to the case in which the integral of $f(x)$ is improper in the sense in which an improper integral has been defined above in the two cases of absolute and of non-absolute convergence.

Let $F(x)$, $\psi(x)$ be two functions which are both continuous in (a, b), and let us suppose that one of the four derivatives $DF(x)$ is finite and equal to the corresponding derivative $D\psi(x)$, at every point of (a, b) with the exception of a set of points G, non-dense in (a, b), and such that the content of the closed set H, obtained by adding to G all its limiting points, is zero. At the points of H, the derivatives $DF(x)$, $D\psi(x)$ may be supposed not to be finite, or not to be equal.

The function $F(x) - \psi(x)$ is then constant throughout any one of the intervals complementary to H; and it has been shewn in § 206 that

$$F(x) - \psi(x)$$

is constant throughout (a, b), in case H be enumerable, but that it need not be constant if H be unenumerable. In the latter case the complementary intervals of H are lines of invariability of $F(x) - \psi(x)$, and the function

$$DF(x) - D\psi(x)$$

is a null-function with an improper integral in (a, b).

If $\phi(x)$ be a continuous function, and one of its four derivatives

$$D\phi(x) = f(x),$$

have a set of points of infinite discontinuity which is enumerable, and $D\phi(x)$ have an improper integral in (a, b), then

$$\int_a^x f(x)\,dx = \phi(x) - \phi(a).$$

For the set of points of infinite discontinuity is non-dense and closed, and has zero content, since $\int_a^x f(x)\,dx$ exists. If $\psi(x) = \int_a^x f(x)\,dx$, the two functions $\psi(x)$, $\phi(x)$ are both continuous, and have the derivatives $D\psi(x)$, $D\phi(x)$ everywhere identical with $f(x)$, except at that set of points, of zero measure, at which $f(x)$ is discontinuous. Hence (see § 224), in any interval containing no points of infinite discontinuity of $f(x)$, the functions $\phi(x)$, $f(x)$ differ by a constant. Since $f(x) - \phi(x)$ has as lines of invariability the intervals complementary to an enumerable closed set, it is constant throughout (a, b); and it is clearly equal to $\phi(a)$.

If the set of points at which $D\phi(x) \equiv f(x)$ is infinite be of the power of the continuum, we can no longer conclude that $\psi(x)$, $\phi(x)$ differ by a constant. In this case we have the theorem:—

If $\phi(x)$ be a continuous function, such that one of its derivatives

$$D\phi(x) \equiv f(x)$$

possesses an improper integral in (a, b), and if the set of points of infinite discontinuity of $D\phi(x)$ be unenumerable, then

$$\int_a^x f(x)\,dx = \phi(x) - \phi(a) + U(x) - U(a)$$

where $U(x)$ is a function with an everywhere-dense set of lines of invariability.

Accordingly, in this case, the fundamental theorem of the Integral Calculus does not hold, in its original form.

The following definition of the definite integral of a function $f(x)$, which in the interval (a, b) possesses an enumerable set G of points of infinite discontinuity, has been given by Hölder*.

Let $F(x)$ be any function which is continuous in (a, b), and is such that, for any two points x_1, x_2, such that no point of G lies in the interval (x_1, x_2), the relation

$$F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)\,dx$$

holds; the function $f(x)$ being assumed to be integrable in every such interval. Then the definite integral of $f(x)$ in any interval (a', b') whatever, contained in (a, b), is defined by $\int_{a'}^{b'} f(x)\,dx = F(b') - F(a')$.

That $F(x)$ is unique, except for an additive constant, has been shewn above. If G were unenumerable, this definition would not suffice to define the integral, because $F(x)$ would not be unique.

GEOMETRICAL INTERPRETATION OF INTEGRATION.

286. Let $f(x)$ be a limited function defined for the interval (a, b), and of which all the values are positive or zero. This function may be considered to define a two-dimensional set of points (x, y) which consists of all the points of which the coordinates satisfy the conditions $a \leqq x \leqq b$, $0 \leqq y \leqq f(x)$. In accordance with Jordan's theory of the measure of sets of points (see § 84), this set has an exterior extent, and an interior extent, and the set of points is measurable when the two have the same value. The extent of a two-dimensional set of points may be regarded as a generalization of the conception of area; thus in the present case, the exterior extent and the interior extent may be spoken of as the *exterior area* and the *interior area* of the space bounded by the axis of x, the two straight lines $x = a$, $x = b$, and the "curve" defined by $y = f(x)$. This set of points G has an area, in the ordinary

* *Math. Annalen*, vol. XXIV, 1884.

sense, when the exterior area and the interior area are equal. The frontier of the two-dimensional set G consists of those points of G which are limiting points of the complementary set $C(G)$, and of those points of $C(G)$ which are limiting points of G. Those points of G which do not belong to the frontier are said to be interior points of G. If a rectangle be drawn on (a, b) as base, and of height greater than the upper limit of $f(x)$ in (a, b), and if this rectangle be divided into a number of rectangular portions by drawing straight lines parallel to the axes of coordinates, then if the number of these rectangles is increased indefinitely, in such a manner that the maximum of the diagonals has the limit zero, then the interior extent of the given two-dimensional set of points is the limit of the sum of those rectangles every point of each of which is an interior point of G. The exterior extent is the limit of the sum of those rectangles, each of which contains at least one point which is either an interior point or a point of the frontier of G.

If, at any stage of the subdivision into rectangles, those sides which are on (a, b) be $\delta_1, \delta_2, \dots \delta_n$, the two sums just referred to are

$$\sum_1^n \delta U(\delta), \quad \sum_1^n \delta L(\delta),$$

where $U(\delta)$, $L(\delta)$ are the upper and lower limits of $f(x)$ in the interval δ; and the limits of these sums are the upper and lower integrals of $f(x)$ in the interval (a, b).

It thus appears that the upper integral $\overline{\int_a^b} f(x)\,dx$ is the exterior extent of the set of points defined by $a \leqq x \leqq b$, $0 \leqq y \leqq f(x)$; and the lower integral $\underline{\int_a^b} f(x)\,dx$ is the interior extent of the same set. If $f(x)$ be integrable, the upper and lower integrals are equal, and the set of points is measurable in accordance with Jordan's definition of measure. Thus the integral represents the area defined as the measure of the set of points, when that set is measurable.

In case $f(x)$ be limited, but not always positive or zero, we may take $f(x) = f_1(x) - f_2(x)$, where $f_1(x) = f(x)$ when $f(x)$ is positive or zero, and $f_1(x) = 0$, when $f(x)$ is negative; with a corresponding definition of $f_2(x)$. In case the two sets of points (x, y) for which $a \leqq x \leqq b$, $0 \leqq y \leqq f_1(x)$ and $a \leqq x \leqq b$, $0 \leqq y \leqq f_2(x)$ are both measurable, the integral $\int_a^b f(x)\,dx$ is the excess of the measure of the first of the two sets over that of the second; and this may be interpreted as the excess of that part of the area defined by $x = a$, $x = b$, $y = 0$, $y = f(x)$ which is above the x-axis over that part which is below it.

If the two sets of points be not measurable, the exterior and interior extents of the first set are $\overline{\int_a^b} f_1(x)\,dx$, $\underline{\int_a^b} f_1(x)\,dx$; and those of the second set are $\overline{\int_a^b} f_2(x)\,dx$, $\underline{\int_a^b} f_2(x)\,dx$ respectively.

The upper integral $\overline{\int_a^b} f(x)\,dx$ is then the excess of the exterior extent of the set $a \leqq x \leqq b$, $0 \leqq y \leqq f_1(x)$, over the interior extent of the set $a \leqq x \leqq b$, $0 \leqq y \leqq f_2(x)$; whilst the lower integral $\underline{\int_a^b} f(x)\,dx$ is the excess of the interior extent of the first of the sets, over the exterior extent of the second set.

The condition of integrability of the function $f(x)$ is that the frontier which consists of the set of points $a \leqq x \leqq b$, $y = f(x)$ when closed by adding the limiting points, shall be a set of zero measure; this measure being that which is applicable to two-dimensional sets. This is the condition for the existence of the area in the ordinary sense of the term, and is equivalent to that of the existence of the corresponding integral.

If a linear set of points G be defined on the x-axis, which is limited and lies in the interval (a, b), then a function $f(x)$ may be defined by the rule that $f(x) = 1$, if x be a point of G, and $f(x) = 0$, if x be not a point of G. This set G has always an exterior extent, and an interior extent, which are given by

$$\overline{\int_a^b} f(x)\,dx, \quad \underline{\int_a^b} f(x)\,dx$$

respectively, as may be seen by referring to the definitions. For it is easily seen that the exterior or the interior linear extent of G is numerically identical with the corresponding extent of the two-dimensional set, defined by the function $y = f(x)$. When G is measurable in accordance with Jordan's definition of a linear measure, the function $f(x)$ is integrable in (a, b), and $\int_a^b f(x)\,dx$ is the measure of G. This measure may be regarded as a generalization of the notion of length of a linear interval. The condition that a linear set G be measurable is that its frontier, which consists of those points of G which are limiting points of $C(G)$, and of those points of $C(G)$ which are limiting points of G, have the linear measure zero.

LEBESGUE'S THEORY OF INTEGRATION.

287. A definition of integration has been developed* by Lebesgue which is applicable to a more extensive class of functions than those which are integrable in accordance with Riemann's definition. The theory depends essentially upon the employment of the conception of the measure of a set of

* See his memoir "Intégrale, Longueur, Aire," *Annali di Mat.*, series III a, vol. VII, 1902; also his *Leçons sur l'intégration*, Paris, 1904.

points, in the sense in which the term is employed by Borel and Lebesgue. It has been shewn in Chapter III, that a set which is measurable in accordance with the definition employed by Jordan is also measurable in accordance with the definition employed by Borel and Lebesgue, but that the converse does not hold.

A function $f(x)$ defined for the interval (a, b), is said to be summable, if the set of points x of the interval (a, b), for which $A < f(x) < B$, is always measurable, whatever numbers A and B may be.

A function $f(x)$ which satisfies the condition stated in the definition may or may not be limited.

The set of points of (a, b) for which $f(x)$ has a fixed value k is measurable, if $f(x)$ is a summable function. For this set is the set of points common to the measurable sets for which $k - \delta < f(x) < k + \delta$, where δ has a sequence of values converging to zero; hence, by a theorem of § 82, the set for which $f(x) = k$, is measurable.

Let $f(x)$ be a summable function which is limited in (a, b), and is never negative. Let the interval (L, U) of variation of $f(x)$ be divided into any n parts $(a_0, a_1), (a_1, a_2), \dots (a_{n-1}, a_n)$, where $a_0 = L, a_n = U$. Let e_ι be the linear set of points in (a, b) for which $f(x) = a_\iota$, and let e_ι' be the linear set of points in (a, b) for which $a_\iota < f(x) < a_{\iota+1}$; let E denote the two-dimensional set of points for which $a \leqq x \leqq b, 0 \leqq y \leqq f(x)$. Those points of the set E for which the values of x belong to e_ι form a two-dimensional set, of which the measure is $a_\iota m(e_\iota)$; and those points of E for which the values of x belong to e_ι' form a set which contains a set of measure $a_\iota m(e_\iota')$, and is itself contained in a set of measure $a_{\iota+1} m(e_\iota')$.

The set E contains a set of measure

$$\sum_{\iota=0}^{n} a_\iota m(e_\iota) + \sum_{\iota=1}^{n} a_{\iota-1} m(e'_{\iota-1}) = M;$$

and it is contained in a set of measure

$$\sum_{\iota=0}^{n} a_\iota m(e_\iota) + \sum_{\iota=1}^{n} a_\iota m(e'_{\iota-1}) = M'.$$

The interior measure of E is $\geqq M$, the measure of the set which E contains; and the exterior measure of E is $\leqq M'$, the measure of the set in which E is contained. The measures M and M' of the two sets differ from one another by not more than $(b - a)\alpha$, where α is the greatest of the numbers $a_\iota - a_{\iota-1}$. If n be increased indefinitely, and the sub-division of (L, U) be such that the greatest interval α converges to zero, we see that the limits of

$$\sum_{\iota=0}^{n} a_\iota m(e_\iota) + \sum_{\iota=1}^{n} a_{\iota-1} m(e'_{\iota-1})$$

$$\sum_{\iota=0}^{n} a_\iota m(e_\iota) + \sum_{\iota=1}^{n} a_\iota m(e'_{\iota-1})$$

both exist, being identical in value, and that E is measurable, its measure being the common limit. It is easily seen, by superimposing two sets of sub-divisions, that the common limit is independent of the particular sets of sub-divisions of (U, L). It has thus been proved that:—

If $f(x)$ be a limited summable function which is never negative in the interval (a, b), then the two-dimensional set E of points (x, y) defined by $a \leqq x \leqq b$, $0 \leqq y \leqq f(x)$ is measurable, and its measure $m(E)$ is the common limit of the sums given above.

The value of the Lebesgue integral of $f(x)$ in the interval (a, b) is defined to be the measure $m(E)$.

It may be shewn that:—

The Lebesgue integral of $f(x)$ lies between the upper and lower integrals of $f(x)$, and is identical with the Riemann integral in case the latter exists.

For it is clear that $m(E)$ is not greater than the sum Σ_m employed in § 252, in defining the upper integral $\overline{\int_a^b} f(x)\,dx$, and hence $m(E) \leqq \overline{\int_a^b} f(x)\,dx$.

In a similar manner it can be shewn that $m(E) \geqq \underline{\int_a^b} f(x)\,dx$.

If G be any measurable set of points contained in (a, b), and e_ι, e_ι' now denote those measurable sets of points of G at which $f(x) = a_\iota$, $a_\iota < f(x) < a_{\iota+1}$ respectively, then the common limit of the two sets of numbers M, M', formed as before, defines *the Lebesgue integral of the summable function $f(x)$ relatively to the measurable set G.*

288. If $f(x)$ be limited and summable, but be not restricted to be positive, we can express $f(x)$ as the difference of two summable functions $f_1(x)$, $f_2(x)$ each of which is positive or zero. Thus $f_1(x) = f(x)$, when $f(x) \geqq 0$, and $f_1(x) = 0$, when $f(x) < 0$; also $f_2(x) = -f(x)$, when $f(x) \leqq 0$, and $f_2(x) = 0$, when $f(x) > 0$. The two sets of points $a \leqq x \leqq b$, $0 \leqq y \leqq f_1(x)$, and $a \leqq x \leqq b$, $0 \leqq y \leqq f_2(x)$, being measurable, their measures may be denoted by $m(E_1)$ and $m(E_2)$. A similar statement holds when the interval (a, b) is replaced by a measurable set G contained in that interval.

The Lebesgue integral of a limited summable function $f(x)$ is defined to be $m(E_1) - m(E_2)$, where E_1, E_2 are the two sets of points above defined.

The measures of the two-dimensional sets E_1, E_2 are the areas in the extended sense of the term, which are defined by the parts of the function $f(x)$ which are respectively above and below the axis of x.

The measures $m(E_1)$, $m(E_2)$ are identical with Jordan's measures of the same sets, in case the latter measures exist; and thus Lebesgue's value of the integral is in agreement with the value according to Riemann's definition, when the latter is applicable. Sets of points which are not measurable according to Jordan's system are in general measurable in accordance with the Borel-Lebesgue definition; accordingly functions which are not integrable according to Riemann's definition may be so according to Lebesgue's definition. It is not known whether sets exist which are not measurable; but, as we have seen in Chapter III, all the sets which are defined in the various modes usually employed, are measurable. Thus Lebesgue's definition of a definite integral has the advantage over that of Riemann, in that all summable limited functions are integrable in accordance with it.

The essential distinction between the two definitions of an integral as the limit of a sum is that, in Riemann's definition, a system of successive sub-divisions of the interval (a, b) of the variable is taken as the basis, whereas in Lebesgue's definition, a system of successive sub-divisions of the interval (L, U) of the function is fundamental.

The relation of Lebesgue's integral to the fundamental theorem of the Integral Calculus, and to the problems which arise in connection therewith will be dealt with in Chap. VI. It will there be shewn that if the limited function $f(x)$ possess a limited differential coefficient $f(x)$ in the interval (a, b), then $f'(x)$ is always integrable in accordance with Lebesgue's definition, and $\int_a^x f'(x)\,dx = f(x) - f(a)$. It has been seen in § 264 that this does not always hold when the definition of Riemann is employed.

289. If $b > a$, $\int_b^a f(x)\,dx$ may be defined to be $-\int_a^b f(x)\,dx$. It is clear that if $f(x)$ is integrable in (a, b) it is integrable in any part (a, x) of (a, b), and that

$$\int_a^b f(x)\,dx = \int_a^x f(x)\,dx + \int_x^b f(x)\,dx.$$

If $\phi_1, \phi_2, \dots \phi_m$ be limited summable functions, and if

$$F(\phi_1, \phi_2, \dots \phi_m) = \chi(x),$$

be a function which is continuous with respect to $(\phi_1, \phi_2, \dots \phi_m)$, and is limited in the interval (a, b), it is also a summable function.

If L_1, U_1 are the lower and upper limits of ϕ_1 in the interval (a, b), we may divide the interval (L_1, U_1) into parts (L_1, y_1), (y_1, y_2), $\dots (y_{n-2}, y_{n-1})$, (y_{n-1}, U_1), each of which is less than a prescribed number ϵ. Let $\psi_1(x)$ be defined as follows:—$\psi_1(x) = L_1$, for all values of x such that $L_1 \leqq \phi_1(x) < y_1$; $\psi_1(x) = y_1$, for all values of x such that $y_1 \leqq \phi_1(x) < y_2$; and generally $\psi_1(x) = y_r$ for all values of x such that $y_r \leqq \phi_1(x) < y_{r+1}$. It is clear that the function $\psi_1(x)$ is summable, and it is such that $|\phi_1(x) - \psi_1(x)| < \epsilon$. Let the functions $\psi_2(x)$, $\psi_3(x)$, $\dots \psi_m(x)$ which correspond to $\phi_2(x)$, $\phi_3(x)$, $\dots \phi_m(x)$ respectively, be defined in a similar manner. Since F is a continuous function of $\phi_1, \phi_2, \dots \phi_m$, therefore, corresponding to a fixed positive number η, a number ϵ can be found such that

$$|F(\phi_1, \phi_2, \dots \phi_m) - F(\psi_1, \psi_2, \dots \psi_m)| < \eta,$$

when $\psi_1, \psi_2, \dots \psi_m$ are defined as above for the value of ϵ which corresponds to η. Thus if $F(\psi_1, \psi_2, \dots \psi_m)$ be denoted by $\lambda(x)$, we have

$$|\chi(x) - \lambda(x)| < \eta:$$

also, if L, U be the lower and upper limits of $\chi(x)$, in (a, b), we have

$$L-\eta < \lambda(x) < U+\eta,$$

for the whole interval (a, b). Now let η have successively the values of the numbers in a decreasing sequence $\eta_1, \eta_2, \ldots \eta_n \ldots$ which converges to the limit zero, and let $\lambda_1(x), \lambda_2(x), \ldots \lambda_n(x), \ldots$ be the corresponding functions $\lambda(x)$. If A, B be any two numbers in the interval (L, U), the sets of points for which $A < \lambda_1(x) < B$, $A < \lambda_2(x) < B$, &c., are all measurable; and the set of points for which $A < \chi(x) < B$ is such that each point belongs to an infinite number of the sets for which $A < \lambda(x) < B$, and is therefore, by a theorem of § 82, measurable. It thus appears that the function $\chi(x)$ is summable, and therefore has an integral in accordance with Lebesgue's definition.

In particular, *the sum, or the product, of a finite number of summable functions is summable.*

If $f_1(x)$, $f_2(x)$ *be summable in the interval* (a, b), *then*

$$\int_a^b [f_1(x)+f_2(x)]\,dx = \int_a^b f_1(x)\,dx + \int_a^b f_2(x)\,dx.$$

It having been shewn above that $f_1(x)+f_2(x)$ is a summable function, we observe that if $\psi_1(x)$, $\psi_2(x)$ be two functions defined as above, each of which has only a finite number of values in the interval (a, b), and be such that

$$|f_1(x)-\psi_1(x)| < \epsilon, \quad |f_2(x)-\psi_2(x)| < \epsilon,$$

then $$\int_a^b [f_1(x)+f_2(x)]\,dx$$

differs from $$\int_a^b [\psi_1(x)+\psi_2(x)]\,dx$$

by less than $2\epsilon(b-a)$.

Also $$\int_a^b [\psi_1(x)+\psi_2(x)]\,dx = \int_a^b \psi_1(x)\,dx + \int_a^b \psi_2(x)\,dx,$$

each integral being an integral in accordance with Riemann's definition;

hence since $$\int_a^b \psi_1(x)\,dx$$

differs from $$\int_a^b \phi_1(x)\,dx$$

by less than $\epsilon(b-a)$,

and $$\int_a^b \psi_2(x)\,dx$$

differs from $$\int_a^b \phi_2(x)\,dx$$

by less than $\epsilon(b-a)$,

it follows that $$\int_a^b [f_1(x)+f_2(x)]\,dx$$

differs from
$$\int_a^b f_1(x)\,dx + \int_a^b f_2(x)\,dx$$
by less than $4\epsilon(b-a)$. Since ϵ is arbitrarily small, the equality contained in the theorem is established.

That
$$\int_a^x f(x)\,dx,$$
for a function $f(x)$ which is summable in (a, b), is a continuous function of x, is established in the same manner as in the case of a function integrable in accordance with Riemann's definition.

290. A general theory of integration has been developed by W. H. Young independently* of the work of Lebesgue, in two memoirs. In the second of these memoirs, the theory there developed is brought into relation with the work of Lebesgue. The domain of the independent variable, for which the function is defined, is taken to be any set of points, and the following definition of integration of a function with regard to such a set is formulated:—

Let the fundamental set be divided into measurable components in any conceivable way, and let the measure of each component be multiplied by the upper (lower) limit of the values of the function at points of that component, and the sum of all such products be formed; then the outer (inner) measure of the integral is defined to be the lower (upper) limit of all such summations.

If it be assumed either (1) that all sets are measurable, or (2) that all functions are summable, then the outer and inner measures of the integral are equal to one another, and their common value defines the integral of the function with respect to the fundamental set of points.

291. Lebesgue has extended his definition so as to afford a definition of an absolutely convergent improper integral. It is clearly sufficient to take the case of an unlimited summable function $f(x)$ which is nowhere negative in the interval (a, b) for which it is defined.

Let $a_0, a_1, a_2, \dots a_n, \dots$ be a sequence of increasing numbers, such that $a_0 = 0$, and that a_n has no upper limit as the index n is indefinitely increased: also let the differences $a_1 - a_0, a_2 - a_1, \dots a_{n+1} - a_n, \dots$ be limited, having η as their upper limit. Consider the two series
$$\sigma = \sum_{r=0}^{\infty} a_r m(e_r) + \sum_{r=0}^{\infty} a_r m(e_r'),$$
$$\sigma' = \sum_{r=0}^{\infty} a_r m(e_r) + \sum_{r=0}^{\infty} a_{r+1} m(e_r'),$$
where e_r, e_r' have the same meaning as in § 287.

* See his papers "On upper and lower integration," *Proc. Lond. Math. Soc.*, ser. 2, vol. II; also "On the general theory of integration," *Phil. Trans.*, vol. CCIV, 1905.

Since the difference of the two series is $\sum_{r=0}^{\infty} (a_{r+1} - a_r) m(e_r')$, which is less than $\eta \sum_{r=0}^{\infty} m(e_r')$, it is clear that the two series are either both convergent, or are both divergent. Let us suppose that the function $f(x)$ is such that both series are convergent; it can then be easily shewn that they are still convergent when further numbers are interpolated between each consecutive pair of the numbers $a_0, a_1, a_2, \ldots$, and the corresponding new series are formed; for by this process σ is increased, and σ' is diminished. Therefore as the process of further sub-division of the interval $(0, \infty)$ proceeds, in any manner consistent with the continual diminution of η to the limit zero, the sums σ, σ' both converge to one and the same number. By superimposition of different systems of sub-division it can also be directly shewn that the limit to which σ and σ' converge is independent of the particular system of sub-division chosen. The common limit of σ and σ', when it exists, is then defined to be the value of the improper integral $\int_a^b f(x)\,dx$. In order that an improper integral of the function $f(x)$ may exist, it is necessary, though not sufficient, that $f(x)$ be a summable function; and also that the measure of those points (x) at which $f(x)$ is $\geqq$ an arbitrarily great number N shall be arbitrarily small for a sufficiently great value of N. For it is a necessary consequence of the convergence of the above series, that $\sum_{r=n}^{\infty} \{m(e_r) + m(e_r')\}$, which is the measure of that set of points at which $f(x) \geqq a_n$, should have a value which converges to zero, as n and a_n are indefinitely increased. It is however not necessary that the content of the set K_∞ of all the points of infinite discontinuity should be zero; in fact it is even possible that the improper integral may exist whilst every point of (a, b) is a point of infinite discontinuity.

It will now be shewn* that Lebesgue's definition can be replaced by one which differs from that of de la Vallée-Poussin only in the one respect, that the convergent sequence of proper integrals $\int f_n(x)\,dx$ consists of Lebesgue integrals, which are not necessarily Riemann integrals.

From the condition of convergence of the second series, it follows that, corresponding to an arbitrarily chosen positive number ϵ, we may determine s so that

$$\sigma' = \sum_{r=0}^{r=s} a_r m(e_r) + \sum_{r=0}^{r=s-1} a_{r+1} m(e_r') + R,$$

where $R < \epsilon$; whilst at the same time η may be chosen so small that σ'

* Hobson, *Proc. Lond. Math. Soc.* ser. 2, vol. IV, p. 144.

differs from $\int_a^b f(x)\,dx$ by less than ϵ. Now let $a_s = N$, and let $f_n(x)$ be that function which $= f(x)$, for $f(x) < N$, and $= N$, for $f(x) \geqq N$.

The Lebesgue proper integral $\int_a^b f_n(x)\,dx$ is then the limit, when η converges to zero, of the sum

$$\sum_{r=0}^{r=s} a_r m(e_r) + \sum_{r=0}^{r=s-1} a_{r+1} m(e_r') + a_s \sum_{r=s+1}^{\infty} m(e_r) + a_s \sum_{r=s}^{\infty} m(e_r');$$

and this sum is equal to

$$\sum_{r=0}^{r=s} a_r m(e_r) + \sum_{r=0}^{s-1} a_{r+1} m(e_r') + S,$$

where $S < R < \epsilon$. Keeping $a_r = N$ fixed, we may now, if necessary, diminish η by interpolating further numbers between the pairs of numbers $a_0, a_1, a_2, \ldots$, until we have the new sum which corresponds to

$$\sum_{r=0}^{r=s} a_r m(e_r) + \sum_{r=0}^{s-1} m(e_r') + S,$$

differing from $\int_a^b f_n(x)\,dx$ by less than ϵ; the part S not having been increased by any diminution of η. We thus find that σ' differs from $\int_a^b f_n(x)\,dx$ by less than ϵ, when N is sufficiently great, and η sufficiently small. Also σ' has been taken to differ from $\int_a^b f(x)\,dx$ by less than ϵ, η having been chosen sufficiently small. Since ϵ is arbitrarily small, it is clear that $\int_a^b f_n(x)\,dx$ converges to $\int_a^b f(x)\,dx$, as N is increased indefinitely.

It has now been shewn that de la Vallée-Poussin's definition of an improper integral may be extended to the case in which the integrals $\int_a^b f_n(x)\,dx$ exist only in the sense defined by Lebesgue. This definition is then equivalent to that of Lebesgue. It is clear that Harnack's definition is only capable of extension, in the case in which K_∞ has zero content. If the condition, that K_∞ have zero content, be satisfied, the reasoning in § 274 is applicable without essential change; and in that case Harnack's definition of an improper integral can be extended to the case in which the proper integrals employed in that definition exist only in the sense defined by Lebesgue. Thus, in this case, all three definitions are equivalent to one another.

INTEGRALS WITH INFINITE LIMITS.

292. The definition of the integral of a limited integrable function given in § 251 is applicable only to the case in which both the limits a, b are definite points, and in which therefore the interval of integration is finite.

Let $x_1, x_2, \dots x_n, \dots$ be a sequence of increasing numbers which has no upper limit; it may then happen that the sequence of integrals

$$\int_a^{x_1} f(x)\,dx,\quad \int_a^{x_2} f(x)\,dx, \dots \int_a^{x_n} f(x)\,dx, \dots$$

has a definite limit A, independent of the particular sequence $\{x_n\}$ chosen. When this is the case $f(x)$ is said to have an integral

$$\int_a^{\infty} f(x)\,dx,$$

in the unlimited interval (a, ∞), the value of this integral being A. It has been presupposed that, in every interval (a, x), the function $f(x)$ is limited and integrable.

If the integrals

$$\int_{x_1}^{b} f(x)\,dx,\quad \int_{x_2}^{b} f(x)\,dx, \dots \int_{x_n}^{b} f(x)\,dx, \dots$$

where $x_1, x_2, \dots x_n \dots$ is a sequence of descending values of x which has no lower limit, all exist, and have a limit B independent of the particular sequence chosen, the limit B is denoted by

$$\int_{-\infty}^{b} f(x)\,dx.$$

If the two integrals

$$\int_0^{\infty} f(x)\,dx,\quad \int_{-\infty}^{0} f(x)\,dx,$$

as thus defined, both exist, their sum is denoted by

$$\int_{-\infty}^{\infty} f(x)\,dx.$$

The three numbers

$$\int_a^{\infty} f(x)\,dx,\quad \int_{-\infty}^{b} f(x)\,dx,\quad \int_{-\infty}^{x} f(x)\,dx,$$

being the limits of integrals, and not themselves in the proper sense of the term integrals, belong to the class of improper integrals.

In each case it is necessary, but not sufficient, for the existence of these improper integrals, that $f(x)$ be integrable in every finite interval contained in the intervals (a, ∞), $(-\infty, b)$ or $(-\infty, \infty)$; and it will be at first assumed that $f(x)$ is limited in every such finite interval, and thus has therein a proper integral.

In case the integral $\int_{-c}^{c} f(x)\,dx$ have a definite limit, as c is indefinitely increased, that limit is said to be the *principal value* of

$$\int_{-\infty}^{\infty} f(x)\,dx.$$

This principal value may exist, even when the integral

$$\int_{-\infty}^{\infty} f(x)\,dx,$$

as defined above, does not exist; but in case the latter do exist, its value is equal to its principal value.

The necessary and sufficient condition for the existence of the integral

$$\int_{a}^{\infty} f(x)\,dx,$$

is that (1) *the integral exist in every interval* (a, x) *where* $x > a$, *and* (2) *that, corresponding to every arbitrarily chosen positive number* ϵ, *a value* ξ *of* x *can be found such that*

$$\left| \int_{\xi}^{\xi'} f(x)\,dx \right| < \epsilon,$$

for every value of ξ' *such that* $\xi' > \xi$.

A similar condition applies to the case of

$$\int_{-\infty}^{b} f(x)\,dx.$$

293. It was shewn in § 253 that the necessary and sufficient condition that $f(x)$ be integrable in the interval (a, b) is that, if $\delta_1^{(m)}, \delta_2^{(m)}, \ldots \delta_{n_m}^{(m)}$ be a particular set of intervals whose sum is (a, b), and of which Δ_m is the greatest, $\sum_1^{n_m} D\delta$ shall converge to zero, as m is increased without limit, the system being subject to the condition that Δ_m have the limit zero; the fluctuation of $f(x)$ in δ being denoted by D. If this condition be satisfied for any one such succession of sub-divisions, then it is satisfied for every other such system. We have to enquire how far a corresponding condition applies to the case of an integral through an infinite interval.

Since $\int_{a}^{\infty} f(x)\,dx$, when it exists, is the limit

$$\lim_{b=\infty} \int_{a}^{b} f(x)\,dx,$$

we see that

$$\int_{a}^{\infty} f(x)\,dx$$

is given as the repeated limit $\lim\limits_{b=\infty} \lim\limits_{\Delta=0} U\delta$. The question then arises whether, and under what conditions, the order of the repeated limits may be reversed, so that

$$\int_a^\infty f(x)\,dx = \lim_{\Delta=0} \overset{\infty}{\Sigma} U\delta = \lim_{\Delta=0} \overset{\infty}{\Sigma} L\delta$$

where $\overset{\infty}{\Sigma} U\delta$ or $\overset{\infty}{\Sigma} L\delta$ denotes the limit of the sum of the $U\delta$, or the $L\delta$, taken through a finite number of intervals δ, as the number of these intervals is increased indefinitely owing to continual increase in b.

In the first place, it is clearly necessary for the truth of this proposition that the limits $\overset{\infty}{\Sigma} U\delta$, $\overset{\infty}{\Sigma} L\delta$ should exist, and that their difference $\overset{\infty}{\Sigma} D\delta$ should converge to zero when Δ does so.

Let us consider a sequence of intervals (a, x_1), (x_1, x_2), ... (x_{n-1}, x_n) ..., where x_n has no upper limit as n is increased indefinitely. Assuming that $f(x)$ has a proper integral in each of these intervals, a system of successive sub-divisions of (a, x_1) can be found such that $\Sigma D\delta$ converges to zero as the greatest of the δ does so; and thus a set of intervals δ exists in (a, x_1) such that

$$\Sigma D\delta < \tfrac{1}{2}\epsilon,$$

where ϵ is a fixed arbitrarily chosen number. Similarly a set of intervals can be chosen for (x_1, x_2) such that $\Sigma D\delta < \frac{1}{2^2}\epsilon$, and so on; thus for (x_{n-1}, x_n) a set of sub-divisions can be found such that $\Sigma D\delta < \frac{1}{2^n}\epsilon$. A set of sub-divisions can accordingly be found for the unlimited interval (a, ∞), such that $\Sigma D\delta$ for (a, x_n) is $< \epsilon\left(1 - \frac{1}{2^n}\right)$. The sum $\Sigma D\delta$ thus found converges, as n is indefinitely increased, to a value which is $< \epsilon$. By taking a sequence of values of ϵ, converging to zero, we now see that a system of successive sub-divisions of (a, ∞) can be found such that $\overset{\infty}{\Sigma} D\delta$ converges to zero, as the greatest of the δ converges to zero. Conversely, if a system of sub-divisions of (a, ∞) exist such that $\overset{\infty}{\Sigma} D\delta$ converges to zero as the greatest of the δ does so, it follows that $f(x)$ is integrable in any finite interval contained in (a, ∞). For let the successive sub-division be so far advanced that $\overset{\infty}{\Sigma} D\delta < \eta$, where η is some fixed number, and consider any interval (α, β) in (a, ∞). Then $\Sigma D\delta$ taken for those δ, finite in number, which either lie wholly inside (α, β), or have one end inside (α, β), is less than η; thus we have a set of sub-divisions of (α, β) such that $\Sigma D\delta < \eta$. By letting η decrease through a sequence of values which converges to zero, we see that a system of successive sub-divisions of (α, β) exists, such that $\Sigma D\delta$ converges to zero, and hence that $f(x)$ is integrable in (α, β).

It has thus been established that *the necessary and sufficient condition that a limited function $f(x)$ defined for (a, ∞) be integrable in every finite* *interval contained in* (a, ∞) *is that a system of successive sub-divisions of* (a, ∞) *should exist such that* $\sum^{\infty} D\delta$ *converges to a value* Σ, *which itself converges to zero as the greatest of the intervals* δ *converges to zero.*

The convergence of $\sum^{\infty} D\delta$ to zero, for a particular system of successive sub-divisions, is not sufficient to ensure the existence of the improper integral

$$\int_a^{\infty} f(x)\, dx,$$

but only that

$$\int_a^b f(x)\, dx$$

shall exist for every value of b. In this respect an integral through an infinite interval differs from one through a finite interval; since, in the latter case, the convergence of the finite sum $\Sigma D\delta$ to zero, for a particular system of sub-divisions, is sufficient to ensure the existence of the integral.

294. In the case of a finite interval of integration, the convergence of $\Sigma U\delta$, $\Sigma L\delta$, where U, L are the upper and lower limits of the function in the interval δ, to one and the same definite limit, follows as a consequence of the convergence of $\Sigma D\delta$ to zero; but in the case of the integral through (a, ∞), the convergence of $\sum^{\infty} U\delta$, $\sum^{\infty} L\delta$ does not necessarily follow from that of $\sum^{\infty} D\delta$.

It will however be shewn that, *if for a function* $f(x)$, *limited in the interval* (a, ∞), *a system of successive sub-divisions of* (a, ∞) *exist, such that* $\sum^{\infty} U\delta$, $\sum^{\infty} L\delta$ *have the same definite value* Σ *which converges to a definite number* A, *as the greatest interval* δ *converges to zero, then the integral* $\int_a^{\infty} f(x)\, dx$ *exists, and its value is* A.

From the existence, and convergence to the same value, of $\sum^{\infty} U\delta$, $\sum^{\infty} L\delta$, the convergence of $\sum^{\infty} D\delta$ to zero follows, and therefore $f(x)$ is integrable in any finite interval of (a, ∞).

Again $\sum_1^n U\delta - \int_a^{a+\sum_1^n \delta} f(x)\, dx \leqq \sum_1^n D\delta$, for any finite value of n; and hence, if all the δ's are so small that $\sum^{\infty} D\delta < \eta$, we have

$$\sum_1^n U\delta - \int_a^{a+\sum_1^n \delta} f(x)\, dx < \eta.$$

Since $\sum_1^n U\delta$ converges, as n is increased indefinitely, to a definite value Σ, a number m can be found such that, if $n \geqq m$, $\sum_{n+1}^{\infty} U\delta < \eta$; hence

$$\left| \sum^{\infty} U\delta - \int_a^{a+\sum_1^n \delta} f(x)\, dx \right| < 2\eta, \text{ if } n \geqq m.$$

Now we may suppose the successive sub-division to be so far advanced that $\left| \overset{\infty}{\Sigma} U\delta - A \right| < \eta$; we then have

$$\left| A - \int_a^{a+\sum_1^n \delta} f(x)\, dx \right| < 3\eta, \text{ for } n \geqq m.$$

It follows that $\left| A - \int_a^{\xi} f(x)\, dx \right| < 3\eta$, for all values of ξ such that $\xi = a + \sum_1^n \delta$, $n \geqq m$. If then X be any number greater than the least of these values of ξ, two values of ξ exist between which X lies; and if ξ_1 be the smaller of these, we have $\int_{\xi_1}^{X} f(x)\, dx < U\sigma$, where σ is the greatest of all the intervals δ, and U is the upper limit of $f(x)$ in (a, ∞). It now follows that

$$\left| A - \int_a^X f(x)\, dx \right| < 3\eta + U\sigma$$

for all values of X which exceed a fixed number. Since η and σ are both arbitrarily small, it follows that $\int_a^X f(x)\, dx$ converges to the limit A, as X is indefinitely increased, and thus that the integral $\int_a^\infty f(x)\, dx$ exists.

295. It will now be proved that, *if $\int_a^\infty f(x)\, dx$ have a definite finite value, a particular system of successive sub-divisions of the unlimited interval (a, ∞) can always be found, such that $\overset{\infty}{\Sigma} U\delta$ exists for each set of the system, and converges to the value of the integral as the greatest of the intervals δ converges to zero.*

For in (a, x_1) we can find a set of intervals such that

$$\Sigma U\delta - \int_a^{x_1} f(x)\, dx < \tfrac{1}{2}\epsilon;$$

in (x_1, x_2) we can find a set such that $\Sigma U\delta - \int_{x_1}^{x_2} f(x)\, dx < \frac{1}{2^2}\epsilon$; and generally in (x_{n-1}, x_n) a set can be found such that $\Sigma U\delta - \int_{x_{n-1}}^{x_n} f(x)\, dx < \frac{1}{2^n}\epsilon$. Thus a set exists in (a, x_n) such that $\overset{x_n}{\Sigma} U\delta - \int_a^{x_n} f(x)\, dx < \epsilon\left(1 - \frac{1}{2^n}\right)$. Now x_n can be taken so great that $\left| \int_a^{x_n} f(x)\, dx - \int_a^\infty f(x)\, dx \right| < \eta$.

Thus an infinite set $\overset{\infty}{\Sigma} U\delta$ exists such that

$$\overset{\infty}{\Sigma} U\delta - \overset{x_n}{\Sigma} U\delta < \frac{1}{2^n}\epsilon,$$

and that

$$\left| \overset{\infty}{\Sigma} U\delta - \int_a^\infty f(x)\, dx \right| < \eta + \epsilon.$$

Therefore $\Sigma U\delta$ converges to a definite limit $\overset{\infty}{\Sigma} U\delta$ which differs from $\int_a^\infty f(x)$ by less than $\eta + \epsilon$, or by not more than ϵ, since η is arbitrarily small. By taking a sequence of values of ϵ, we obtain a sequence of sub-divisions of (a, ∞) such that $\overset{\infty}{\Sigma} U\delta$ converges to the value $\int_a^\infty f(x)\,dx$.

It may also be possible to define a system of successive sub-divisions, such that the greatest of the intervals converges to zero as the successive sub-division advances, but such that $\overset{x_n}{\Sigma} U\delta$ does not converge to a value $\overset{\infty}{\Sigma} U\delta$; and thus the theorem

$$\int_a^\infty f(x)\,dx = \lim_{\Delta=0} \overset{\infty}{\Sigma} U\delta = \lim_{\Delta=0} \overset{\infty}{\Sigma} L\delta$$

only holds provided the sub-divisions of (a, ∞) be such that $\overset{\infty}{\Sigma} U\delta$ exists for each successive set of intervals, after some fixed one.

It can also be shewn that *when the integral* $\int_a^\infty f(x)\,dx$ *has a definite finite value, then, for every system of successive sub-divisions of* (a, ∞) *which is such that* $\overset{\infty}{\Sigma} D\delta$ *exists and converges to zero as the greatest* δ *converges to zero, the set of numbers* $\overset{\infty}{\Sigma} U\delta$ *exists and converges to the value of the integral.*

For since, for every value of n,

$$\overset{n}{\Sigma} U\delta \leqq \int_a^{a+\overset{n}{\Sigma}\delta} f(x)\,dx + \overset{n}{\Sigma} D\delta,$$

we see that $\overset{n}{\Sigma} U\delta$ converges, as n is indefinitely increased, to the value

$$\int_a^\infty f(x)\,dx + \overset{\infty}{\Sigma} D\delta;$$

and hence, since $\overset{\infty}{\Sigma} D\delta$ has the limit zero, as the successive sub-division proceeds indefinitely, it follows that $\overset{\infty}{\Sigma} U\delta$ has the limit $\int_a^\infty f(x)\,dx$. For a system of successive sub-divisions which is not such that $\overset{\infty}{\Sigma} D\delta$ converges to zero, $\overset{n}{\Sigma} U\delta$ does not in general converge to a definite finite value as n is increased indefinitely.

296. The definitions of $\int_a^\infty f(x)\,dx$, $\int_{-\infty}^b f(x)\,dx$ may be extended to the case in which $f(x)$ has points of infinite discontinuity. If the improper integral $\int_a^X f(x)\,dx$ exists for every value of X which is $> a$, and if it converges to a definite limit as X increases indefinitely through a sequence of values, independently of the particular sequence, then that limit defines

$$\int_a^\infty f(x)\,dx.$$

The integrals $\int_x^\infty f(x)\,dx$, $\int_{-\infty}^x f(x)\,dx$, when they exist, possess many of the properties of a proper or improper integral $\int_a^x f(x)\,dx$. These integrals are continuous functions of the finite limit x. For

$$\int_{-\infty}^x f(x)\,dx = \int_{-\infty}^a f(x)\,dx + \int_a^x f(x)\,dx,$$

where $a < x$; and since $\int_a^x f(x)\,dx$ is a continuous function of x, so also is $\int_{-\infty}^x f(x)\,dx$.

If a function $F(x)$ be such that the sequence $F(x_1)$, $F(x_2)$, ... $F(x_n)$, ..., where $x_1, x_2, \ldots x_n, \ldots$ is a sequence of values of x having no upper limit, has a definite limit, independent of the particular sequence $\{x_n\}$, then that limit may be denoted by $F(\infty)$, and the function $F(x)$ is said to be continuous at $x = \infty$.

When the integral $\int_a^\infty f(x)\,dx$ exists, the integral $\int_a^x f(x)\,dx$ is continuous for all values of x in the interval (a, ∞), including $x = \infty$.

If the integral $\int_a^x f(x)\,dx$ exist for every finite value of x in the interval (a, ∞), and if $\phi(x)$ be a function which is finite and continuous for every such value of x, and be such that

$$\phi(x) - \phi(a) = \int_a^x f(x)\,dx,$$

then, provided $\phi(x)$ be continuous for $x = \infty$, the function $f(x)$ is integrable in (a, ∞), and $\phi(\infty) - \phi(a) = \int_a^\infty f(x)\,dx$. If the function $\phi(x)$ have a derivative, say $D^+\phi(x)$, which is integrable in every interval (a, x) of (a, ∞); then if the integral be a proper one, or be such an improper one that the relation

$$\phi(x) - \phi(a) = \int_a^x D^+\phi(x)\,dx$$

subsists, then, provided the limit $\phi(\infty)$ exist, we have also

$$\phi(\infty) - \phi(a) = \int_a^\infty D^+\phi(x)\,dx.$$

A similar statement applies to each of the other derivatives of $\phi(x)$. In the case in which $\phi(x) - \phi(a)$ differs from $\int_a^x D^+f(x)\,dx$ by an integrable null-function, this holds also for the limit $x = \infty$.

297. An integral $\int_a^\infty f(x)\,dx$ is said to be *absolutely convergent* when the integral $\int_a^\infty |f(x)|\,dx$ exists; otherwise it is said to be *conditionally* or *relatively convergent.* If $\int_a^\infty |f(x)|\,dx$ exists, then also $\int_a^\infty f(x)\,dx$ exists; for $\left|\int_{x_1}^{x_2} f(x)\,dx\right| \leqq \int_{x_1}^{x_2} |f(x)|\,dx$, and hence the convergence of the latter integral follows from that of the former one.

If $f(x)$ and $f(x)\,\phi(x)$ be both integrable in every interval (a, x) contained in (a, ∞), and if $\int_a^\infty f(x)\,dx$ be absolutely convergent, and if $\phi(x)$ be, from and after some fixed value of x, numerically less than some fixed number, then the integral $\int_a^\infty f(x)\,\phi(x)\,dx$ exists, and is absolutely convergent.*

For since $\int_a^\infty f(x)\,dx$ is absolutely convergent, we can find, corresponding to a fixed positive number σ, a number $\xi > a$, such that $\int_\xi^{\xi+h} |f(x)|\,dx < \sigma$, for all positive values of h; we have then

$$\left|\int_\xi^{\xi+h} f(x)\,\phi(x)\,dx\right| \leqq K\int_\xi^{\xi+h} |f(x)|\,dx \leqq K\sigma,$$

where K is the upper limit of $|\phi(x)|$, and is by hypothesis finite. It is thus seen that $\int_a^\infty f(x)\,\phi(x)\,dx$ is convergent. Also since

$$\int_\xi^{\xi+h} |f(x)\,\phi(x)|\,dx \leqq K\int_\xi^{\xi+h} |f(x)|\,dx \leqq K\sigma,$$

we see that the convergence is absolute.

298. *If $f(x)$ and $f(x)\,\phi(x)$ be both integrable in every interval (a, x), and if $\int_a^\infty f(x)\,dx$ exist, and if further, from and after some fixed value of x, $\phi(x)$ be monotone, and $|f(x)|$, $|\phi(x)|$ be each less than some fixed number, then $\int_a^\infty f(x)\,\phi(x)\,dx$ has a definite finite value.*

If σ be a fixed positive number, a value ξ of x can be so chosen that $\int_\xi^{\xi+h} f(x)\,dx < \sigma$, for every positive value of h; also

$$\int_\xi^{\xi+h} f(x)\,\phi(x)\,dx = \phi(\xi)\int_\xi^{\xi+\theta h} f(x)\,dx + \phi(\xi+h)\int_{\xi+\theta h}^{\xi+h} f(x)\,dx,$$

where θ is in the interval $(0, 1)$. If $|\phi(x)| < K$, for every value of x concerned, we have

$$\left|\int_\xi^{\xi+h} f(x)\,\phi(x)\,dx\right| < 2\sigma K;$$

hence, since σ is arbitrarily small, the integral $\int_a^\infty f(x)\,\phi(x)\,dx$ is convergent.

* Riemann's *Werke*, p. 229; also Pringsheim, *Math. Annalen*, vol. xxxvii, p. 591.

299. An important set of tests of the absolute convergence of an integral $\int_a^\infty f(x)\,dx$ is the following:—

If $f(x)$ be integrable in every interval (a, x), then $\int_a^\infty f(x)\,dx$ converges to a definite finite value, provided $f(x)$ converge to zero, as x is increased indefinitely, in such a manner that one of the expressions $f(x) \cdot x^{1+k}$, $f(x)\,x\,(\log x)^{1+k}$, $f(x)\,x \log x\,(\log\log x)^{1+k}$, ... $f(x)\,x\log x \cdot \log\log x$ ...... $(\log\log \ldots \log x)^{1+k}$, converges to zero as x is indefinitely increased, k denoting some fixed number greater than zero.

The integral $\int_a^\infty f(x)\,dx$ is not convergent, in case $f(x)$ be of invariable sign, from and after some fixed value of x, and provided also any one of the above expressions do not converge to zero as x is increased indefinitely, when k has the value zero.

We see that in the first case, $\int_X^{X+h} |f(x)|\,dx$ is numerically less than one of the expressions

$$C\int_X^{X+h}\frac{dx}{x^{1+k}},\quad C\int_X^{X+h}\frac{dx}{x(\log x)^{1+k}},\quad C\int_X^{X+h}\frac{dx}{x\log x(\log\log x)^{1+k}},\ \ldots$$

where C is a constant dependent on X, which converges to zero as X is indefinitely increased. These expressions have the values

$$\frac{C}{k}\left[\frac{1}{X^k}-\frac{1}{(X+h)^k}\right],\quad \frac{C}{k}\left[\frac{1}{(\log X)^k}-\frac{1}{(\log X+h)^k}\right],$$
$$\frac{C}{k}\left[\frac{1}{(\log\log X)^k}-\frac{1}{(\log\log X+h)^k}\right];$$

hence, k being positive, it is clear that X may be so chosen that

$$\int_X^{X+h}|f(x)|\,dx$$

is less than an arbitrarily fixed number, and thus

$$\int_a^\infty |f(x)|\,dx$$

is convergent. In the second case, k being now zero, we see that

$$\int_X^{X+h} f(x)\,dx$$

is numerically greater than one of the expressions

$$C\int_X^{X+h}\frac{dx}{x},\ C\int_X^{X+h}\frac{dx}{x\log x},\ C\int_X^{X+h}\frac{dx}{x\log x\log\log x},\ \ldots$$

or than one of

$$C\log\frac{X+h}{X},\quad C\log\frac{\log(X+h)}{\log X},\quad C\log\log\frac{\log(X+h)}{\log X},\ \ldots$$

and these expressions increase indefinitely as h is increased. It follows that $\int_a^\infty f(x)\,dx$ is in this case divergent.

300. An important case of convergent integrals, which do not necessarily converge absolutely, is that of the integrals

$$\int_0^\infty \phi(x)\sin x\,dx,\quad \int_0^\infty \phi(x)\cos x\,dx,$$

when $\phi(x)$ is monotone, from and after some fixed value of x, and converges to zero as x is indefinitely increased.

We have $\int_{x_1}^{x_2}\phi(x)\sin x\,dx = \phi(x_1)\int_{x_1}^{\xi}\sin x\,dx + \phi(x_2)\int_{\xi}^{x_2}\sin x\,dx,$

where x_1 is so great that $\phi(x)$ is monotone for $x \geqq x_1$, and $x_2 > x_1$. From this we have

$$\left|\int_{x_1}^{x_2}\phi(x)\sin x\,dx\right| \leqq 2\,|\phi(x_1)| + 2\,|\phi(x_2)| \leqq 4\,|\phi(x_1)|;$$

and hence, if x_1 be taken sufficiently great, $\left|\int_{x_1}^{x_2}\phi(x)\sin x\,dx\right|$ is, for all values of x_2, less than an arbitrarily fixed number ϵ. The convergence of the integral has thus been established. The case of the second integral may be treated in the same manner.

INTEGRATION BY PARTS.

301. Let $f(x)$, $\phi(x)$ be two functions which are continuous in the interval (a, b); also let $Df(x)$ be one of the four derivatives of $f(x)$, and $D\phi(x)$ one of the four derivatives of $\phi(x)$. If the two derivatives $Df(x)$, $D\phi(x)$ be both limited integrable functions in the interval (a, b), the relation

$$\int_a^b f(x)\,D\phi(x)\,dx = \Big[f(x)\,\phi(x)\Big]_a^b - \int_a^b \phi(x)\,Df(x)\,dx$$

is satisfied, where

$$\Big[f(x)\,\phi(x)\Big]_a^b \text{ denotes } f(b)\,\phi(b) - f(a)\,\phi(a).$$

If U be a function which differs from $Df(x)$ only by an integrable null-function, and V differ from $D\phi(x)$ only by an integrable null-function, the above formula may be written in the form

$$\int_a^b V\left(\int_a^x U dx\right)dx = \left[\left(\int_a^x U dx\right)\left(\int_\beta^x V dx\right)\right]_a^b - \int_a^b U\left(\int_\beta^x V dx\right)dx,$$

where α, β are arbitrarily fixed points in the interval (a, b). This general theorem in integration was first* obtained by P. Du Bois Reymond, and is a generalization of Leibnitz's formula for integration by parts,

$$\int_a^b u\frac{dv}{dx}\,dx = \Big[uv\Big]_a^b - \int_a^b v\frac{du}{dx}\,dx.$$

To prove the theorem, let (a, b) be divided into n sub-intervals $\delta_1, \delta_2, \ldots \delta_n$; then, if x_{r-1}, x_r denote the end-points of the sub-interval δ_r, where $x_1 = a$, $x_n = b$, we have

$$\Big[f(x)\,\phi(x)\Big]_a^b = \sum_{r=1}^{r=n}\Big[\phi(x_r)\{f(x_r) - f(x_{r-1})\} + f(x_{r-1})\{\phi(x_r) - \phi(x_{r-1})\}\Big].$$

Now $\dfrac{f(x_r) - f(x_{r-1})}{\delta_r}$ lies between the upper and lower limits of $Df(x)$ in the sub-interval δ_r, and similarly $\dfrac{\phi(x_r) - \phi(x_{r-1})}{\delta_r}$ lies between the upper and lower limits of $D\phi(x)$ in the same interval; therefore it follows that the sum on the right-hand side of the above identity may be written in the form

$$\sum_{r=1}^{r=n}\delta_r[\phi(x_r)\,\chi_r + f(x_{r-1})\,\psi_r],$$

where χ_r is between the upper and lower limits of $Df(x)$, and ψ_r between those of $D\phi(x)$ in the interval δ_r.

Let $\chi_r = Df(x_r) + \epsilon_r$, $\psi_r = D\phi(x_{r-1}) + \zeta_r$, where $|\epsilon_r|$ cannot exceed the fluctuation of $Df(x)$ in the interval δ_r, and $|\zeta_r|$ cannot exceed that of $D\phi(x)$ in the same interval.

We have now

$$\Big[f(x)\,\phi(x)\Big]_a^b = \sum_{r=1}^{r=n}\delta_r[\phi(x_r)\,Df(x_r) + f(x_{r-1})\,D\phi(x_{r-1})] + \sum_{r=1}^{r=n}\delta_r[\phi(x_r)\,\epsilon_r + f(x_{r-1})\,\zeta_r];$$

and the absolute value of the second sum on the right-hand side cannot exceed

$$\Phi\sum_{r=1}^{r=n}\Big\{\delta_r|\epsilon_r|\Big\} + F\sum_{r=1}^{r=n}\Big\{\delta_r|\zeta_r|\Big\},$$

where Φ, F are the upper limits of $|\phi(x)|$, $|f(x)|$ in the interval (a, b). Since $Df(x)$, $D\phi(x)$ are by hypothesis integrable in (a, b), the set of sub-intervals may be so chosen that

$$\sum_{r=1}^{r=n}\Big\{\delta_r|\epsilon_r|\Big\},\quad \sum_{r=1}^{r=n}\Big\{\delta_r|\zeta_r|\Big\}$$

are arbitrarily small. Since $\phi(x)\,Df(x)$, $f(x)\,D\phi(x)$ are by hypothesis integrable, we see that, if the number n be made to increase indefinitely, and

* *Abhandlungen d. Münch. Akad.*, vol. XII, p. 129.

the intervals δ be so chosen that the greatest of them converges to the limit zero, the above identity assures us that

$$\left[f(x)\,\phi(x)\right]_a^b = \int_a^b \phi(x)\,Df(x)\,dx + \int_a^b f(x)\,D\phi(x)\,dx;$$

and thus the theorem is established.

302. The theorem may be extended to the case in which $Df(x)$, $D\phi(x)$ are not restricted to be both limited functions, but may have points of infinite discontinuity forming an enumerable non-dense closed set. It has been shewn above that, in any interval (c, x) contained in (a, b), in which $\phi(x)\,Df(x) + f(x)\,D\phi(x)$ has a proper integral,

$$\left[f(x)\,\phi(x)\right]_c^x = \int_c^x \left\{\phi(x)\,Df(x) + f(x)\,D\phi(x)\right\} dx.$$

It follows from the theorem of § 285, that

$$\left[f(x)\,\phi(x)\right]_a^b = \int_a^b \left\{\phi(x)\,Df(x) + f(x)\,D\phi(x)\right\} dx,$$

provided that the improper integral on the right-hand side exists. If now the two functions $\phi(x)\,Df(x)$, $f(x)\,D\phi(x)$ possess absolutely convergent improper integrals in (a, b), it follows that their sum has the same property, and that

$$\left[f(x)\,\phi(x)\right]_a^b = \int_a^b \phi(x)\,Df(x)\,dx + \int_a^b f(x)\,D\phi(x)\,dx.$$

The theorem has now been proved under the suppositions that $f(x)$, $\phi(x)$ are continuous in (a, b), that $Df(x)$, $D\phi(x)$ have at most points of infinite discontinuity forming a closed enumerable set, and that the functions $f(x)\,D\phi(x)$, $\phi(x)\,Df(x)$ possess absolutely convergent improper integrals in (a, b).

If $f(x)\,\phi(x)$ have a finite limit for $x = \infty$, and be continuous at $x = \infty$, in the sense defined in § 296; and if further the conditions be satisfied, that the functions $f(x)\,D\phi(x)$, $\phi(x)\,Df(x)$ are integrable in the infinite interval (a, ∞), then the formula

$$\left[f(x)\,\phi(x)\right]_a^\infty = \int_a^\infty \phi(x)\,Df(x)\,dx + \int_a^\infty f(x)\,D\phi(x)\,dx$$

holds. If the formula for integration by parts holds for (a, b), whatever finite value b may have, but if $f(x)\,\phi(x)$ be not finite and continuous for $x = \infty$, then one at least of the integrals

$$\int_a^\infty \phi(x)\,Df(x)\,dx, \quad \int_a^\infty f(x)\,D\phi(x)\,dx$$

is infinite, or does not converge to a definite value.

CHANGE OF THE VARIABLE IN AN INTEGRAL.

303. Let $f(x)$ be a limited function, integrable in the interval (a, b). We now assume that x is a continuous function $\psi(y)$ of another variable y, defined for the interval (α, β), where $a = \psi(\alpha)$, $b = \psi(\beta)$, and that $\psi(y)$ is monotone in the interval (α, β). If we further assume that $\psi(y)$ has no lines of invariability in the interval (α, β), then y can be regarded as a single-valued function $\phi(x)$ of x, defined for the interval (a, b) of x; and also $f(x)$ can be regarded as a function $f\{\psi(y)\}$ of y, defined for the interval (α, β).

If $D\psi(y)$ be one of the derivatives of $\psi(y)$, then, provided $D\psi(y)$ be limited and integrable in (α, β), the function $f\{\psi(y)\}D\psi(y)$ is integrable in (α, β), and

$$\int_a^b f(x)\,dx = \int_\alpha^\beta f\{\psi(y)\}\,D\psi(y)\,dy.$$

This result is a generalization of the well-known formula of substitution

$$\int_a^b f(x)\,dx = \int_\alpha^\beta f\{\psi(y)\}\,\psi'(y)\,dy,$$

the particular case of the theorem which arises when $\psi(y)$ is differentiable throughout the interval.

Let the interval (α, β) be divided into n parts $\delta_1', \delta_2', \ldots \delta_n'$; then the interval (a, b) is divided into corresponding parts $\delta_1, \delta_2, \ldots \delta_n$.

Since $\psi(y)$ is a continuous function of y, the intervals $\{\delta'\}$ can be so chosen that in each of them the fluctuation of $\psi(y)$ is less than a fixed arbitrarily chosen positive number ϵ, and hence so that each of the intervals δ is less than ϵ. If then a sequence of diminishing values of ϵ be taken, which converges to the limit zero, we obtain a convergent system of sets $\{\delta'\}$, and corresponding to them a convergent system of sets $\{\delta\}$ of sub-intervals of (a, b). We may assume that $f(x)$ is positive; for if it is not so throughout the interval (a, b), it may be made so by the addition of a properly chosen constant c: then, if $[c + f\{\psi(y)\}]\,D\psi(y)$ be integrable in (α, β), it follows, since $D\psi(y)$ is integrable in that interval, that $f\{\psi(y)\}\,D\psi(y)$ has the same property. Let U_r, L_r denote the upper and lower limits of $f(x)$ in δ_r, or of $f\{\psi(y)\}$ in δ_r'; and let u_r, l_r denote the upper and lower limits of $D\psi(y)$ in δ_r'. The fluctuation D_r' of $f\{\psi(y)\}\,D\psi(y)$ in δ_r' is $\leqq U_r u_r - L_r l_r$, or $\leqq U_r(u_r - l_r) + l_r(U_r - L_r)$. Hence, since $l_r\delta_r' \leqq \delta_r$, it being assumed that $D\psi(y)$ is always positive, or zero, so that l_r is not negative, we have

$$D_r'\,\delta_r' \leqq U(u_r - l_r)\,\delta_r' + \delta_r(U_r - L_r),$$

where U is the upper limit of $f(x)$ in (a, b).

We have now

$$\sum_1^n D_r'\delta_r' \leqq U\sum_1^n (u_r - l_r)\,\delta_r' + \sum_1^n (U_r - L_r)\,\delta_r;$$

and it follows, from the conditions of integrability of $D\psi(y)$ in (α, β), and of $f(x)$ in (a, b), that $f\{\psi(y)\} D\psi(y)$ is integrable in (α, β).

Next, we have

$$\sum_1^n \delta_r f(x_r) - \sum_1^n \delta_r' f\{\psi(y_r)\} D\psi(y_r) = \sum_1^n \delta_r' f\{\psi(y_r)\} \{\lambda_r - D\psi(y_r)\},$$

where λ_r lies between the upper and lower limits of $D\psi(y)$ in the interval δ_r'. The absolute value of the expression on the left-hand side is consequently less than $U\sum_1^n \delta_r'(u_r - l_r)$, which is arbitrarily small, on account of the condition of integrability of $D\psi(y)$ in (α, β). It has thus been established, that

$$\int_a^b f(x)\, dx = \int_\alpha^\beta f\{\psi(y)\} D\psi(y)\, dy.$$

304. The theorem can be extended to the case in which $D\psi(y)$ has points of infinite discontinuity forming an enumerable closed set, provided that $D\psi(y)$ have an absolutely convergent improper integral in the interval (α, β).

For in this case, the equality

$$\int_c^x f(x)\, dx = \int_{c'}^y f\{\psi(y)\} D\psi(y)\, dy$$

holds for any interval (c', y) which contains none of the points of infinite discontinuity of $D\psi(y)$. The integral $\int_c^x f(x)\, dx$ is a continuous function of the upper limit x, and the integral $\int_{c'}^y f\{\psi(y)\} D\psi(y)\, dy$ is a continuous function $F(y) - F(c')$ of the upper limit y. Since $f\{f(y)\}$ is limited in the interval (α, β), $f\{\psi(y)\} D\psi(y)$ has an improper integral in that interval; and the relation $F(y) - F(c') = \int_c^y f\{\psi(y)\} D\psi(y)\, dy$ holds in intervals (c', y), which contain points of infinite discontinuity of $D\psi(y)$. It thus appears that the substitution formula holds for the interval (a, b) together with the corresponding interval (α, β). Precisely similar considerations suffice to shew that the theorem still holds when $f(x)$ has a set of points of infinite discontinuity forming an enumerable closed set, provided $f(x)$ have an absolutely convergent improper integral in (a, b).

305. The theorem of substitution is still valid when $\psi(y)$ is no longer monotone in the interval (α, β), but has a finite number of maxima and minima, or, under certain restrictions, an infinite number of maxima and minima. The function $\psi(y)$ may also have lines of invariability. In this case x has a single value for each value of y in (α, β) defined by the continuous function $\psi(y)$, but the inverse function $\phi(x)$ is not single-valued.

It may happen that the values of x given by $\psi(y)$ do not all lie within the interval (a, b). In this case it is convenient to conceive the function $f(x)$ to be defined for all such values of x, so that it is integrable in every interval of x, such interval extending so far as necessary beyond (a, b). The result of the substitution will consist of two or more integrals with respect to y.

Let us assume that the points y at which $\psi(y)$ has a maximum or minimum, or is the end-point of a line of invariability, form a set of points in (α, β) with the content zero; they can then all be enclosed in a finite set of intervals of which the sum is less than an arbitrarily chosen number ϵ. If it be assumed that $D\psi(y)$ has finite upper and lower limits in (α, β), then the sum of those intervals in (a, b), produced if necessary, which correspond to the finite set of intervals constructed in (α, β), is less than ϵ multiplied by the upper limit of $|D\psi(y)|$ in (α, β); and this may be made as small as we please by choosing ϵ small enough. To each of the remaining intervals of (α, β), when the finite set of intervals is removed, the theorem of § 303 is applicable, it being remembered that along a line of invariability $D\psi(y) = 0$.

In such an interval (α_r, β_r) we have

$$\int_{\alpha_r}^{\beta_r} f\{\psi(y)\}\, D\psi(y)\, dy = \int_{a_r}^{b_r} f(x)\, dx,$$

where (a_r, b_r) corresponds to (α_r, β_r), since $\psi(y)$ is monotone in the interval.

The sum of the integrals on the left-hand side, taken for all the finite number of values of r, differs from

$$\int_{\alpha}^{\beta} f\{\psi(y)\}\, D\psi(y)\, dy$$

by less than ϵ multiplied by the upper limit of $|f\{\psi(y)\}\, D\psi(y)|$ in the interval (α, β); and the sum of the integrals on the right-hand side, taken for all the values of r, differs from $\int_a^b f(x)\, dx$ by less than ϵ multiplied by the upper limit of $|D\psi(y)|$ in (α, β), and by the upper limit of $|f(x)|$ for all the values of x for which $f(x)$ has been defined. Since ϵ is arbitrarily small, the theorem has been proved for the case in which $\psi(y)$ has maxima and minima at a set of points of zero content, and may also have lines of invariability of which the end-points form a set with zero content.

The theorem here established can be extended, as in § 304, to the case in which $D\psi(y)$ has points of infinite discontinuity forming an enumerable closed set, provided $D\psi(y)$ have an improper integral in (α, β). An extension can be made to the case in which $f(x)$ has a set of points of infinite discontinuity, with zero content, on the assumption that $f(x)$ possesses an improper integral in (a, b).

306. The theorem of § 303 may be generalised so as to apply to the case when the limited function $f(x)$ is not necessarily integrable in the interval (a, b). It being assumed as before that $\psi(y)$ is monotone in (α, β), and has no lines of invariability, and that $D\psi(y)$ is limited and integrable in (α, β), it will be shewn that the upper integral of $f(x)$ in (a, b) is equal to the upper integral of $f\{\psi(y)\}\,D\psi(y)$ in (α, β).

Denoting by u_r' the upper limit of $f\{\psi(y)\}\,D\psi(y)$ in the interval δ_r', we see that u_r' lies between $U_r u_r$ and $U_r l_r$. Also δ_r lies between $u_r\delta_r'$ and $l_r\delta_r'$, thus we may write $\delta_r = \delta_r'\{u_r - \theta_r(u_r - l_r)\}$, where $0 \leqq \theta_r \leqq 1$. We have to prove that $\Sigma U_r\delta_r$ and $\Sigma u_r'\delta_r'$ converge to the same limit. Writing $\Sigma U_r\delta_r - \Sigma u_r'\delta_r'$ in the form

$$\Sigma\delta_r'[U_r\{u_r - \theta_r(u_r - l_r)\} - u_r'],$$

or
$$\Sigma\delta_r'(U_r u_r - u_r') - \Sigma\delta_r'\theta_r \,.\, U_r(u_r - l_r),$$

we have
$$\Sigma\delta_r'(U_r u_r - u_r') \leqq \Sigma\delta_r' U_r(u_r - l_r) \leqq U\Sigma\delta_r'(u_r - l_r),$$

and also
$$\Sigma\delta_r'\theta_r U_r(u_r - l_r) \leqq \Sigma\delta_r' U_r(u_r - l_r) \leqq U\Sigma\delta_r'(u_r - l_r).$$

Since $D\psi(y)$ is integrable in (α, β), it follows that $\Sigma\delta_r'(u_r - l_r)$ is arbitrarily small, when the number of intervals δ' is increased; it thus appears that $\Sigma U_r\delta_r - \Sigma u_r'\delta_r'$ is arbitrarily small, and thus $\Sigma U_r\delta_r$, $\Sigma u_r'\delta_r'$ converge to the same limit. It has now been established that:—

If $D\psi(y)$ be one of the derivatives of $\psi(y)$, and if $D\psi(y)$ be limited and integrable in (α, β), the continuous function $\psi(y)$ being monotone and without lines of invariability in (α, β), then

$$\overline{\int_a^b} f(x)\,dx = \overline{\int_\alpha^\beta} f\{\psi(y)\}\,D\psi(y)\,dy.$$

The corresponding theorem for the lower integrals may be established in a similar manner.

It follows from these theorems that, under the conditions stated, if either of the integrals $\int_a^b f(x)\,dx$, $\int_\alpha^\beta f\{\psi(y)\}\,D\psi(y)\,dy$ have a definite value, then the other one has the same definite value. The method of substitution may accordingly be applied to $\int_a^b f(x)\,dx$ without assuming that this integral has a definite value, and may thus be employed to decide the question of the integrability of $f(x)$.

The extensions in § 304 and § 305 may be applied to the case of the upper or the lower integrals of $f(x)$.

307. It may happen that, when x and y are connected by the relation $x = \psi(y)$, an infinite interval for y corresponds to the interval (a, b) of x. For example, $y = \infty$ may correspond to $x = b$: in that case we assume that

$\psi(\infty)$ is continuous, *i.e.* that it is the limit of $\psi(y)$ when y is indefinitely increased. If the conditions of the theorems in §§ 303–305 be satisfied for every interval $(a, b-\epsilon)$ of x, with the corresponding interval (α, β') of y, we have

$$\int_a^{b-\epsilon} f(x)\,dx = \int_\alpha^{\beta'} f\{\psi(y)\}\,D\psi(y)\,dy;$$

and since this holds for every ϵ, we have, on proceeding to the limit $\epsilon = 0$,

$$\int_a^b f(x)\,dx = \int_\alpha^\infty f\{\psi(y)\}\,D\psi(y)\,dy,$$

in accordance with the definition of the improper integral on the right-hand side. Similar considerations apply to the case in which $y = -\infty$ corresponds to $x = a$.

308. If one or both of the limits a, b be indefinitely great, it may happen that finite values of y correspond to $x = a$, $x = b$, or else that one, or both, of the limits of y may be infinite.

The method of procedure adopted above suffices to establish the theorem that, *if $f(x)$ be integrable in the finite or infinite interval (a, b), then the integral $\int_a^b f(x)\,dx$ can be transformed by means of $x = \psi(y)$ into the integral $\int_\alpha^\beta f\{\psi(y)\}\,D\psi(y)\,dy$, in which the interval of integration is finite or infinite, provided that $\psi(y)$ be finite and continuous in (α, β), or have at most in every part of (α, β) points of infinite discontinuity forming a set of zero content; with the further conditions that $D\psi(y)$ be integrable in (α, β), and that in every part of this interval its maxima and minima and end-points of lines of invariability form a set of points of zero content.*

It will be observed that a large class of those improper integrals which have an infinite interval of integration may be transformed into proper integrals. It has been suggested by Kronecker that every such integral through an infinite interval can be transformed into a proper integral.

309. If it be desired to transform the integral $\int_a^b f(x)\,dx$, by means of the relation $y = \phi(x)$, where $\phi(x)$ is a single-valued function of x, then, unless $\phi(x)$ be monotone in the interval (a, b), the inverse function $\psi(y)$ will not be everywhere single-valued.

If it be assumed that $\phi(x)$ is monotone in (a, b), and that $\alpha = \phi(a)$, $\beta = \phi(b)$, a derivative $D\phi(x)$ of $\phi(x)$ is reciprocal to a derivative $D\psi(y)$ of $\psi(y)$. If it be assumed that $\dfrac{1}{D\phi(x)}$ is integrable in (α, β), and that the same holds for $\dfrac{f(x)}{D\phi(x)}$ considered as a function of y, or else that $f(x)$ is integrable in (a, b), then we may use the transformation

$$\int_a^b f(x)\,dx = \int_\alpha^\beta \left\{\frac{f(x)}{D\phi(x)}\right\}_{y=\phi(x)} dy.$$

If $\phi(x)$ be not monotone, $\int_a^b f(x)\,dx$ can not in general be transformed into a single integral in y. If, for example, $\phi(x)$ increases from $x=a$ to $x=k$, and then diminishes from $x=k$ to $x=b$, we must take

$$\int_a^b f(x)\,dx = \int_\alpha^{\phi(k)} \left\{\frac{f(x)}{D\phi(x)}\right\}_{y=\phi(x)} dy + \int_{\phi(k)}^\beta \left\{\frac{f(x)}{D\phi(x)}\right\}_{y=\phi(x)} dy,$$

and the integrals on the right-hand side cannot in general be amalgamated into one integral through the interval (α, β), because in the two integrals the integrand has different values for the same value of y.

Thus, for example, if $y = \sin x$,

$$\int_0^\pi f(x)\,dx = \int_0^1 \left[\frac{f(x)}{\cos x}\right] dy + \int_1^0 \left[\frac{f(x)}{\cos x}\right] dy$$

$$= \int_0^1 \frac{f(\sin^{-1}y)}{\sqrt{1-y^2}}\,dy + \int_0^1 \frac{f\left(\frac{\pi}{2} + \cos^{-1}y\right)}{\sqrt{1-y^2}}\,dy,$$

the value of $\cos x$ in the second integral on the right-hand side of the first equation being negative, and the values of $\sin^{-1}y$, $\cos^{-1}y$ being in the interval $(0, \frac{1}{2}\pi)$.

DOUBLE INTEGRATION.

310. Let G denote a set of points in two-dimensional space, entirely contained in a rectangle with sides parallel to the x and y axes; the set G is accordingly a bounded set. It has been pointed out in § 286 that, if the fundamental rectangle be divided into any number of parts by means of straight lines parallel to the sides, and the sum of those rectangles be taken each point of which belongs to G, then the sum of the rectangles has a definite limit S_1, the interior extent of G, when their number is increased indefinitely in any manner, subject to the condition that the diagonal of the greatest of the rectangles have the limit zero. Also the sum of those rectangles each of which contains at least one point, either of G or of the frontier of G, has, under a similar condition, a definite limit S_2, the exterior extent of G. When $S_1 = S_2$, the set of points G is measurable in accordance with Jordan's definition, and consequently also in accordance with the definition of Borel and Lebesgue; and the set G then has a single definite extent or area, the measure of the set.

Let a function $f(x, y)$ be defined for the bounded set G, which we shall assume to have a definite extent in the sense just explained. We further assume that $|f(x, y)|$ has a definite upper limit for the set G, so that

$f(x, y)$ is a limited function. It will be convenient to assume that $f(x, y)$ is extended to the whole rectangle in which G is contained, by providing that $f(x, y)$ vanishes at every point of the rectangle which belongs to the complementary set $C(G)$.

The definition of the double integral of $f(x, y)$, with respect to the set G, is now similar to Riemann's definition, in § 251, of a single integral of a function with respect to a linear interval.

Let the fundamental rectangle be divided into n_1 rectangular portions $\delta_1^{(1)}, \delta_2^{(1)}, \dots \delta_{n_1}^{(1)}$, so that $\sum_{r=1}^{n} \delta_r^{(1)} = A$, the area of the fundamental rectangle; and let Δ_1 denote the greatest of the diagonals of the rectangular portions. Let these rectangles be further sub-divided in any manner, so that the whole area A is divided into n_2 parts, the greatest diagonal being Δ_2; the rectangular portions of A now being $\delta_1^{(2)}, \delta_2^{(2)}, \dots \delta_{n_2}^{(2)}$. Let this process of sub-division of A be carried on indefinitely, so that at any stage of the process A is divided into n_m rectangular portions $\delta_1^{(m)}, \delta_2^{(m)}, \dots \delta_{n_m}^{(m)}$, the greatest of the diagonals of these portions being Δ_m. If this system of sub-division of A be made in any manner whatever, which is such that the sequence $\Delta_1, \Delta_2, \dots \Delta_m, \dots$ has the limit zero, it will be spoken of as a convergent system of sub-divisions of the area A.

Let $M(\delta_s^{(m)})$ denote any number whatever which is so chosen as to be not greater than the upper limit, and not less than the lower limit, of $f(x, y)$ in $\delta_s^{(m)}$; and consider the sums

$$S_1 = \delta_1^{(1)} M(\delta_1^{(1)}) + \delta_2^{(1)} M(\delta_2^{(1)}) + \dots + \delta_{n_1}^{(1)} M(\delta_{n_1}^{(1)}),$$

$$S_2 = \delta_1^{(2)} M(\delta_1^{(2)}) + \delta_2^{(2)} M(\delta_2^{(2)}) + \dots + \delta_{n_2}^{(2)} M(\delta_{n_2}^{(2)}),$$

....

....

$$S_m = \delta_1^{(m)} M(\delta_1^{(m)}) + \delta_2^{(m)} M(\delta_2^{(m)}) + \dots + \delta_{n_m}^{(m)} M(\delta_{n_m}^{(m)}).$$

If the sequence $S_1, S_2, \dots S_m, \dots$ be convergent, and have the same number S for its limit, whatever convergent system of sub-divisions of A be employed, and however the numbers $M(\delta_s^{(m)})$ be chosen, subject only to their limitation in relation to the upper and lower limits of $f(x, y)$ in the rectangle $\delta_s^{(m)}$, then the function $f(x, y)$ is said to be integrable in the set G, and the number S defines the value of the double integral. This double integral, when it exists, may be denoted by $\int_{(G)} f(x, y)(dxdy)$.*

It will be observed that the double integral has been defined as the single limit of a sum, and accordingly the sign of integration is here employed

* The extension of Riemann's definition to double integrals was given by H. J. S. Smith, *Proc. Lond. Math. Soc.*, vol. VI, p. 152, 1875, and by Thomae, *Einleitung in die Theorie der bestimmten Integrale*, p. 33, 1875; also *Schlömilch's Zeitschrift*, vol. XXI, p. 224.

each multiplied by the corresponding upper limit of the function, is obtained by diminishing Σ_m by less than $U-L$ multiplied by the sum of those of the δ's which are not interior to rectangles ϵ. Those of the rectangles δ which are interior to a rectangle ϵ are also elements of the superimposed system. If we consider any one of the rectangles ϵ, the sum of the areas of those rectangles δ which encroach on this rectangle ϵ, but are not contained in its interior, is less than the perimeter of this rectangle ϵ, multiplied by Δ_m the greatest diagonal of all the rectangles δ. Therefore the sum of the areas of all those rectangles δ which are not entirely interior to some rectangle ϵ is less than $\Delta_m P$, where P is the sum of the perimeters of all the rectangles ϵ. Hence the sum Σ_m is diminished in the new system, obtained by superimposition, by less than $(U-L)\Delta_m P$. Since Δ_m diminishes indefinitely as m is indefinitely increased, we may choose m so great that $(U-L)\Delta_m P$ is less than an arbitrarily chosen positive number η. Hence the sum for the superimposed system is $>\Sigma_m - \eta$, or $>\Sigma - \eta$. But this new sum is certainly $<\Sigma' + \zeta$. Now η can be chosen so small that the conditions that the new sum is $<\Sigma' + \zeta$ and $>\Sigma - \eta$ are incompatible with one another. It thus appears that Σ and Σ' cannot be unequal.

The existence of the lower integral follows from the fact that it is the upper integral of $-f(x, y)$, if its sign be changed.

The condition for the existence of $\int_{(G)} f(x, y)\,(dxdy)$ is then that *the fundamental rectangle can be divided into a number of rectangular parts such that the sum of the products of the area of each part into the fluctuation of $f(x, y)$ in that part is less than an arbitrarily chosen positive number.*

By reasoning precisely similar to that in § 254, the intervals and neighbourhoods being replaced by rectangles, it can be shewn that the condition for the existence of the double integral is reducible to the following form:—

The necessary and sufficient condition for the existence of the double integral of a limited function defined for a domain G contained in a rectangular area is that those points at which the saltus of the function is $\geqq k$, form, for each value of the positive number k, a set of points of zero content.

It will be observed that those points of A which, without being points of G, are points of the frontier of G, will in general be points of discontinuity of the function $f(x, y)$ extended to the whole domain A by attributing zero values to the function for all points of A which do not belong to G. It has, however, been assumed, in assuming that the frontier G has zero measure, that those points of A which are limiting points of G, without belonging to G, form a set with zero measure. Therefore the content of the points at which the saltus of the function is $\geqq k$, is unaffected by the points of the boundary of G which do not belong to it; and, in the above statement of the condition of integrability, the set of points at which the saltus is $\geqq k$ may be taken to be points of G only.

As in the case of single integrals, the necessary and sufficient condition for the existence of the integral may be expressed in the form, that *the set of points of discontinuity of the function in the domain for which it is defined must form a set of which the two-dimensional measure is zero.*

It is clear that the definition, and the condition for the existence, of an integral of a function of any number of variables, called a multiple integral, are of the same character as in the case of a double integral. For simplicity, the investigations will be here restricted to the case of double integrals; and it will then be easy to extend the results to triple or to n-fold integrals.

312. In defining the double integral, successive sub-division of the fundamental rectangle into rectangular portions has been employed; it will however be seen that this mode of sub-division is not essential, but is a special case of a more general mode. Let us consider a closed connex set e of points in the fundamental rectangle. The distance between a pair of points of e has an upper limit, when every possible pair is considered; this upper limit we may speak of as the *diameter* of the closed connex set e. Let the fundamental rectangle be divided into a definite number m of closed connex sets $e_1, e_2, \dots e_r$, the frontier of each of which has zero content; let $U_{e_1}, U_{e_2}, \dots U_{e_r}$ denote the upper limits of the limited function $f(x, y)$ in the various sets e; and let $L_{e_1}, L_{e_2}, \dots L_{e_r}$ denote the corresponding lower limits; also let d denote the greatest of the diameters of the sets e.

Let us now consider an indefinite succession of such sub-divisions of the fundamental rectangle, subject to the condition that, as the number r of the parts of the rectangle is increased indefinitely as the successive sub-division proceeds, the number d converges to zero.

The two sums $\sum_1^r U_e m(e)$, $\sum_1^r L_e m(e)$, where $m(e)$ denotes the measure of the set e, converge to two definite numbers. For $\sum_1^r U_e m(e)$ cannot increase as the successive sub-division proceeds, and it is not less than L multiplied by the area of the fundamental rectangle; and therefore it converges to a definite limit. The proof is precisely similar for the second sum. It will be shewn that these limits are independent of the mode of successive sub-division of the fundamental rectangle, provided the conditions stated above are satisfied. It follows that, in the definition of the double integral, any mode of successive sub-division, of the type specified, may be employed; the sub-division into rectangular parts being merely a special case of the general mode. The division of the rectangle into curvilinear portions by means of two families of ordinary curves is a special case of the mode of sub-division specified above.

In the first place, since $\sum_1^r U_e m(e)$ is greater than a fixed number, it follows that there exists a lower limit $\overline{\Sigma}$ of all the numbers, $\sum_1^r U_e m(e)$, for every value of r, and for every possible system of sub-divisions of the type considered.

It will next be shewn that any system of rectangular sub-divisions, as in § 311, is such that $\overline{\Sigma}$ is the lower limit of the sum Σ_m for such sub-divisions, *i.e.* that $\Sigma = \overline{\Sigma}$.

Consider the rectangles $\delta_1^{(m)}, \delta_2^{(m)}, \dots \delta_{n_m}^{(m)}$, for which $\Sigma_m = \Sigma\delta^{(m)} U(\delta^{(m)})$. We observe that a system of sub-divisions $e_1, e_2, \dots e_s$ may be so chosen that $\Sigma U_e m(e)$ is less than $\overline{\Sigma} + \eta$, where η is an arbitrarily chosen positive number. We may suppose m to be fixed so large that Δ_m is less than the least of the diameters of $e_1, e_2, \dots e_s$. Some of the rectangles δ will be interior to one or other of the e's, and others will contain points of the frontier of two or more of the e's. We may divide $\Sigma\delta^{(m)} U(\delta^{(m)})$ into two parts, corresponding to this distinction relatively to the rectangles δ. Denoting these two parts by Σ_{m1}, Σ_{m2} respectively, we have

$$\Sigma_m = \Sigma_{m1} + \Sigma_{m2} \leqq \Sigma U_e m(e) + (U - L)\Sigma,$$

where Σ denotes the sum of all those of the δ's which are not interior to one of the e's.

Now, since the frontiers of all the e's have zero content, it follows that, when m is sufficiently large, $(U - L)\Sigma$ is less than an arbitrarily chosen number ζ; and therefore $\Sigma_m < \overline{\Sigma} + \eta + \zeta$. Also $\Sigma_m \geqq \overline{\Sigma}$; hence, since η and ζ are arbitrarily small, it follows that Σ, the limit of Σ_m, is equal to $\overline{\Sigma}$.

Lastly, it will be shewn that for any system of successive sub-divisions e, of the type defined above, the limit of $\sum_1^r U_e m(e)$, when the sub-division is continued indefinitely, is the same as for a rectangular system of sub-divisions. The general theorem will then have been established, that, *whatever system of sub-divisions be taken, of the type defined above, the limit of* $\sum_1^r U_e m(e)$ *is one and the same number, which is the upper integral of* $f(x, y)$ *in the fundamental rectangle.*

Taking a fixed set of rectangular sub-divisions, for which the sum $\Sigma\delta^{(m)} U(\delta^{(m)})$ is less than $\Sigma + \eta$, we may suppose the successive sub-divisions into parts e to be so far advanced that d is less than the least of the sides of the rectangular sub-divisions. Some of the parts e will then be interior to a rectangle δ, but none will contain such a rectangle in its interior.

We have then

$$\sum_1^r U_e m(e) \leq \Sigma_m + Pd(U-L),$$

where P denotes the sum of the perimeters of all the rectangles. Since d is arbitrarily small, and $\Sigma_m < \Sigma + \eta$, we see that

$$\sum_1^r U_e m(e) < \Sigma + \zeta,$$

where ζ is arbitrarily small. Hence $\sum_1^r U_e m(e)$ lies between Σ and $\Sigma + \zeta$; and therefore the lower limit of $\sum_1^r U_e m(e)$ is Σ. The corresponding theorems for the lower integral may be deduced by changing the sign of $f(x, y)$.

If we denote the measure of a portion e of the fundamental rectangle by δe, we may denote the double integral of $f(x, y)$ over this rectangle by $\int f(x,y)\,de$, where $dxdy$ is written, as in § 310, for de, in case a rectangular system of sub-divisions of the fundamental rectangle be supposed to be employed.

The definition of the integral of a summable function, due to Lebesgue (§ 287), can be immediately extended to the case of summable functions with two or more variables. In this case the formal theory is exactly similar to that developed in §§ 287, 288, for the case of summable functions of a single variable. The sets e_ι, e_ι' there employed, must be interpreted to be sets of points in two or more dimensions, the measures $m(e_\iota)$, $m(e_\iota')$ denoting two-dimensional, or multi-dimensional, measures.

REPEATED INTEGRALS.

313. The actual evaluation of a double integral over the fundamental rectangle, of which the sides are $x = x_0$, $x = x_1$, $y = y_0$, $y = y_1$, is usually made to depend upon the evaluation of successive single integrals taken first with respect to one of the variables, and then with respect to the other. The expression

$$\int_{x_0}^{x_1} dx \int_{y_0}^{y_1} f(x, y)\,dy,$$

in which $f(x, y)$ is supposed to be integrated first with respect to y, for a constant value of x, and then with respect to x, is called a repeated integral. Similarly, the expression

$$\int_{y_0}^{y_1} dy \int_{x_0}^{x_1} f(x, y)\,dx,$$

in which the integrations are performed in the reverse order, is also called a repeated integral. The question of the existence of these repeated integrals, and in any given case, their relation with one another, and with the double integral, will be here investigated. It will be observed that the double integral has been defined as a single limit; whereas the repeated integrals are each, when they exist, obtained as the results of repeated limits. We have then to investigate whether, or under what conditions, a double integral is capable of representation as a repeated limit of one of the forms indicated. It cannot be assumed *a priori* that the existence of the double integral necessarily implies the existence, for each value of x, of the single integral

$$\int_{y_0}^{y_1} f(x, y)\, dy$$

as a definite number. Neither is the existence of this single integral, as a definite number, necessary for the existence, as a definite number, of the repeated integral

$$\int_{x_0}^{x_1} dx \int_{y_0}^{y_1} f(x, y)\, dy.$$

In fact, if we assume that the upper and lower integrals

$$\overline{\int_{y_0}^{y_1}} f(x, y)\, dy, \quad \underline{\int_{y_0}^{y_1}} f(x, y)\, dy$$

have different values for some of the values of x, it may happen that the two repeated limits

$$\int_{x_0}^{x_1} dx \overline{\int_{y_0}^{y_1}} f(x, y)\, dy, \quad \int_{x_0}^{x_1} dx \underline{\int_{y_0}^{y_1}} f(x, y)\, dy$$

have identical values.

The repeated integral will consequently be regarded as, in this case, existing; and thus it may be defined as

$$\int_{x_0}^{x_1} dx \overline{\underline{\int_{y_0}^{y_1}}} f(x, y)\, dy,$$

where the upper or lower integral with respect to y is to be taken indifferently, provided the repeated limit exists as a definite number.

In a similar manner $\int_{y_0}^{y_1} dy \overline{\underline{\int_{x_0}^{x_1}}} f(x, y)\, dx,$

when it has a definite value independent of whether the upper or lower integral with respect to x be used, will be regarded as the repeated integral, first with respect to x and then with respect to y.

314. It was first established by P. Du Bois Reymond* that, when the limited function $f(x, y)$ has a double integral in the fundamental rectangle, then the two repeated integrals exist and are each equal to the double integral.

* *Crelle's Journal*, vol. xciv, 1883, p. 277.

We shall first give a proof* of this theorem which exhibits its relation with the theory of sets of points.

The following preliminary theorem will be first established :—

If $\int f(x, y)\,(dxdy)$, *taken through the fundamental rectangle, have a definite value, then the set of values of* x *for which the single integral* $\int f(x, y)\,dy$, *taken with* x *constant, has a definite value, defines a set of points on a side of the rectangle, of linear measure equal to the length of that side.*

It follows from this theorem, that the set is everywhere-dense in the interval, and of cardinal number c. Moreover, the points at which

$$\overline{\int} f(x, y)\,dy,\quad \underline{\int} f(x, y)\,dy$$

differ from one another form a set of measure zero.

On the assumption of the existence of the double integral, the set K of all the points at which the saltus of $f(x, y)$ is $\geqq k$, where k is an arbitrarily chosen positive number, is a closed set of plane content zero. If a straight line be drawn parallel to the y-axis through the point x of the side of the rectangle, then the component of K on this straight line will be denoted by K_x, and its linear content by $I(K_x)$. It has been shewn in § 108, that, σ denoting a prescribed positive number, the linear content of that set of points x, on the side of the rectangle, for which $I(K_x) \geqq \sigma$, is zero; and thus that $I(K_x)$, considered as a function of x, is an integrable null-function, for each value of k. The function $\chi(x)$, $= \lim_{k=0} I(K_x)$, is also an integrable null-function; for the set of points at which $\chi(x)$ does not vanish is made up of those sets of points at which $I(K_x^{(1)})$, $I(K_x^{(2)})$, ... $I(K_x^{(n)})$, ... do not vanish, where $K^{(1)}$, $K^{(2)}$, ... $K^{(n)}$, ... correspond to a diminishing sequence of values of k converging to the limit zero: and since each of these sets has zero measure, it follows that the set of points at which $\chi(x)$ does not vanish has zero measure. At any point x_1, at which $\chi(x)$ vanishes, $I(K_{x_1})$ vanishes for every value of k.

It should be observed that, at a point of K_x, it is not necessarily the case that $f(x, y)$, considered as a function of y, with x constant, has its saltus $\geqq k$; in fact, this saltus may be less than k, or may be zero. However, all the points at which the saltus of $f(x, y)$ taken with x constant, is $\geqq k$, are certainly included in the set K_x.

For any fixed value of x, the upper and lower integrals

$$\overline{\int} f(x, y)\,dy,\quad \underline{\int} f(x, y)\,dy$$

* The investigation is founded on that of Schönflies, *Bericht über die Mengenlehre*, p. 193.

both exist, and the two have equal values at any point of the everywhere-dense set of points x, at which $\chi(x)$ vanishes. The preliminary theorem above stated has thus been established.

Let $F(x)$ denote $\int f(x, y)\,dy$, where $F(x)$ consequently has a single determinate value at each point x at which

$$\overline{\int} f(x, y)\,dy, \quad \underline{\int} f(x, y)\,dy$$

are equal. At any point of that set of zero measure, at which

$$\overline{\int} f(x, y)\,dy, \quad \underline{\int} f(x, y)\,dy$$

have different values, $F(x)$ is regarded as indeterminate; and the upper and lower integrals are the upper and lower limits of indeterminacy. It will now be shewn that the function $F(x)$, so defined, is an integrable function.

Let x' be a point on the side of the rectangle, such that the component $K_{x'}$ of K, on the line $x = x'$, has content $< \sigma$. A finite number of intervals $\epsilon_1, \epsilon_2, \ldots \epsilon_m$ can be determined on the straight line $x = x'$, neither abutting on, nor overlapping, one another, such that their sum $\epsilon_1 + \epsilon_2 + \ldots + \epsilon_m > b - \sigma$, where b is the length of the side of the rectangle parallel to the y-axis, and such also as to contain, in their interiors and at their ends, no points at which the saltus of $f(x, y)$ is $\geqq k$. For each point of one of these intervals ϵ there exists a rectangle with the point at the centre, such that the fluctuation of $f(x, y)$ in that rectangle is $< k$. The breadths of these rectangles for all points of ϵ must have a finite minimum, for otherwise there would exist a point of ϵ which would belong to K. It follows that, for the point x', an interval $(x' - \alpha, x' + \beta)$ can be determined, such that the straight lines $x = x' - \alpha$, $x = x' + \beta$, intersect all the rectangles corresponding to all the points of the intervals $\epsilon_1, \epsilon_2, \ldots \epsilon_m$. If x_1, x_2 be any two points in the interval

$$(x' - \alpha,\ x' + \beta),$$

we have
$$|F(x_1) - F(x_2)| < bk + \sigma(U - L).$$

Now a finite number of separate intervals $\delta_1, \delta_2, \ldots \delta_r$ can be determined on the side x of the rectangle (length $= a$), such that $\delta_1 + \delta_2 + \ldots + \delta_r > a - \eta$, where η is a prescribed positive number, and such that each point of each of the intervals δ is a point x', for which an interval $(x' - \alpha, x' + \beta)$ can be determined as above. By applying the Heine-Borel theorem we see that

$$\delta_1, \delta_2, \ldots \delta_r$$

will all be covered by a finite number of these intervals $(x' - \alpha, x' + \beta)$. It thus appears that the x-side of the rectangle can be divided into a finite number of parts

$$\tau_1, \tau_2, \ldots \tau_p,$$

and $\lambda_1, \lambda_2, \ldots \lambda_q$,

such that $\tau_1 + \tau_2 + \ldots + \tau_p > a - \eta$,

and $\lambda_1 + \lambda_2 + \ldots + \lambda_q < \eta$;

and such that the fluctuation of $F(x)$ in any one of the parts τ is

$$< bk + \sigma(U - L).$$

Let $k = \epsilon/2b, \ \sigma = \epsilon/2(U - L)$;

we see then that $F(x)$ is such that the x-side of the fundamental rectangle can be divided into a finite number of parts, such that the sum of those parts, in which the fluctuation of $F(x)$ is $\geqq \epsilon$, is less than the arbitrarily chosen number η. It follows that $F(x)$ is integrable along the side of the rectangle.

It has now been shewn, on the assumption of the existence of the double integral, that the repeated integral

$$\int F(x)\,dx, \quad \text{or} \quad \int dx \overline{\underline{\int}} f(x, y)\,dy,$$

taken through the fundamental rectangle, has a definite value. Moreover, this value is equal to that of the double integral. For, let the fundamental rectangle be divided up by means of straight lines parallel to the y-axis, through the end-points of each interval of the two finite sets $\{\tau\}$ and $\{\lambda\}$. Any one of the rectangles so constructed, with τ as base and with height b, can be divided into parts by means of straight lines parallel to the x-axis, such that, in each one of a number of these parts the sum of whose heights is $> b - \sigma$, the fluctuation of $f(x, y)$ is $< k$. The fundamental rectangle has now been divided into a finite number of parts, such that the sum of the products of each part multiplied by the upper limit of $f(x, y)$ in that part exceeds the sum of the products of each part multiplied by the lower limit of the function in that part, by less than $abk + (a\sigma + b\eta)(U - L)$, which is arbitrarily small.

Also $$\int dx \overline{\underline{\int}} f(x, y)\,dy, \quad \text{and} \quad \int f(x, y)(dx\,dy),$$

both lie between the two sums of products, and therefore differ from one another by less than $abk + (a\sigma + b\eta)(U - L)$. The equality of the double integral and the repeated integral is thus put in evidence by the mode of sub-division of the rectangle which has been adopted.

Similar reasoning applies to the repeated integral in which the integration is taken first with respect to x, and then with respect to y.

It has thus been established that, *if the double integral through the fundamental rectangle exist, then the two repeated integrals also exist, and are each equal to the double integral.*

All the points at which $\chi(x)$ vanishes are points of continuity of the function $F(x)$; but there may also be other points at which $F(x)$ is continuous; because the existence of a saltus of $f(x, y)$ at a point (x, y) is consistent with $f(x, y)$ being continuous with respect to x, and also with respect to y, at the point.

The function $F(x)$ may be replaced by $\psi(x)$, the most nearly continuous function related to it (§ 192). We thus have

$$\int f(x, y)\,(dx\,dy) = \int \psi(x)\,dx.$$

It has been assumed that the set of points G, for which $f(x, y)$ is defined, is measurable in accordance with Jordan's definition of a measurable set; and thus, that the double integral of the limited function may be replaced by that of the function $f(x, y)$, defined for all points of a rectangle which contains G, by the convention that $f(x, y)$ shall vanish at all those points of the rectangle which do not belong to G. If the set G be such that each straight line parallel to the y-axis contains points of G which fill up a finite, or an indefinitely great number of continuous intervals, or more generally, if the set of such points for each value of x be linearly measurable, then the integral

$$\int f(x, y)\,dy,$$

taken along the whole segment of the line between the sides of the rectangle, may be replaced by the same integral taken through the component of G on the same segment. In particular, if the points of G on the straight line through the point x consist of all the points in the linear interval

$$(f_1(x), f_2(x)),$$

we may replace

$$\int f(x, y)\,dy$$

by

$$\int_{f_1(x)}^{f_2(x)} f(x, y)\,dy;$$

and therefore in this case,

$$\int_G f(x, y)(dx\,dy) = \int_{x_0}^{x_1} dx \int_{f_1(x)}^{f_2(x)} f(x, y)\,dy.$$

315. A simple proof* of the fundamental theorem of § 314, will be given, which depends upon the fact that, for any limited function $f(x, y)$, if the operation of taking the upper integral first with respect to y, and then with

* This method of proof was first employed by Harnack; see his edition of Serret's *Differential and Integral Calculus*, p. 282. Other proofs of this kind have been given by Arzelà, *Mem. dell' Ist. di Bologna*, ser. 5, vol. II, p. 123; by Jordan, *Liouville's Journal*, ser. 4, vol. VIII, p. 84, or *Cours d'Analyse*, vol. I, p. 42; also by Pringsheim, *Sitzungsberichte d. Münch. Akad.*, vol. XXVIII, p. 59, and vol. XXIX, p. 39. See also Pierpont's paper "On multiple integrals," *Trans. Amer. Math. Soc.*, vol. VI, 1905, where a proof of this character for multiple integrals is given.

respect to x, be performed, the result cannot exceed the upper double integral; and that, similarly, the result of successively taking the lower integrals with respect to x and to y cannot be less than the lower double integral: thus

$$\underline{\int} f(x, y)(dxdy) \leqq \underline{\int} dx \underline{\int} f(x, y)\, dy \leqq \overline{\int} dx \overline{\int} f(x, y)\, dy$$
$$\leqq \overline{\int} f(x, y)(dxdy),$$

the integrals being all taken over the fundamental rectangle.

If $f(x, y)$ be integrable, so that

$$\overline{\int} f(x, y)(dxdy) = \underline{\int} f(x, y)(dxdy),$$

it follows that

$$\overline{\int} dx \overline{\int} f(x, y)\, dy = \overline{\int} dx \underline{\int} f(x, y)\, dy$$
$$= \underline{\int} dx \overline{\int} f(x, y)\, dy = \underline{\int} dx \underline{\int} f(x, y)\, dy,$$

and thus that the repeated integral

$$\int dx \int f(x, y)\, dy$$

has a definite value equal to the double integral.

To establish the theorem, let the rectangle be divided into a number of parts δ by means of straight lines parallel to the sides. Since the double integral is assumed to exist, this may be done in such a manner that, ϵ denoting an arbitrarily chosen positive number, the conditions

$$\int f(x, y)(dxdy) - \epsilon \leqq \Sigma\, \{\delta L(\delta)\} \leqq \Sigma\, \{\delta U(\delta)\} \leqq \int f(x, y)(dxdy) + \epsilon$$

are satisfied, where the summation Σ is taken for all the rectangles δ, and $U(\delta)$, $L(\delta)$ denote the upper and lower limits of $f(x, y)$ in a rectangle δ. Now, if we take the upper and lower integrals of $f(x, y)$ along a straight line parallel to the y-axis, we have

$$\overline{\int} f(x, y)\, dy \leqq \Sigma_1\, \{\delta' U(\delta)\},$$
$$\underline{\int} f(x, y)\, dy \geqq \Sigma_1\, \{\delta' L(\delta)\},$$

where the summation Σ_1 refers to all those rectangles δ which are intersected by the straight line along which the upper and lower integrals are taken; and in the case when that straight line is along one or more boundaries of the rectangles δ, Σ refers to all the rectangles on one side of that line: also δ'

denotes that interval along the line of integration which is in the rectangle δ. It follows that

$$\overline{\int} dx \overline{\int} f(x, y)\, dy \leqq \Sigma \{\delta U(\delta)\} \leqq \overline{\int} f(x, y)\,(dxdy) + \epsilon$$

and $$\underline{\int} dx \underline{\int} f(x, y)\, dy \geqq \Sigma \{\delta L(\delta)\} \geqq \underline{\int} f(x, y)\,(dxdy) - \epsilon;$$

and since these inequalities hold for every value of ϵ, we have

$$\overline{\int} dx \overline{\int} f(x, y)\, dy \leqq \overline{\int} f(x, y)\,(dxdy)$$

$$\underline{\int} dx \underline{\int} f(x, y)\, dy \geqq \underline{\int} f(x, y)\,(dxdy);$$

and thus the theorem is established.

316. The converse questions now arise whether, from the existence of one of the repeated integrals, or from the existence and equality of both repeated integrals, that of the double integral can be inferred. The answer to both questions must be in the negative. Continuity of a function $f(x, y)$ with respect to x and y separately does not necessarily imply continuity with respect to (x, y); moreover, the saltus of the function at a point with respect to x, when y has a constant value, or with respect to y when x has a constant value, is not necessarily equal to the saltus of the function with respect to (x, y). It may happen that the component of K on a straight line parallel to one of the axes may consist of points some or all of which are points of continuity of the function when considered as a function of one variable on that straight line. Thus K may have a plane content greater than zero; and yet the linear content of the points on all straight lines parallel to the axes, at which the linear saltus of the function is $\geqq k$, may be zero. Hence either, or both, of the repeated integrals may exist, whilst for values of k, the sets K are not of zero content*; and therefore whilst $f(x, y)$ does not admit of a double integral. The relation of Lebesgue integrals with repeated integrals will be considered in Chap. VI.

EXAMPLES.

1.† For the rectangle bounded by $x=0$, $x=1$, $y=0$, $y=1$, let $f(x, y)=1$, for all rational values of x, and $f(x, y)=2y$, for all irrational values of x. We have then $\int_0^1 f(x, y)\, dy = 1$, whatever value x may have; and hence the repeated integral

$$\int_0^1 dx \int_0^1 f(x, y)\, dy$$

* An incorrect theorem relating to this point has been given by Schönflies, see his *Bericht*, p. 197. In this theorem the condition that K should be closed is stated to be the condition for the existence of the double integral. If, however, K were not closed, it could not represent the set of points at which any function had a saltus $\geqq k$. The examples given by Schönflies do not in reality accord with his theorem.

† Thomae, *Schlömilch's Zeitschrift*, vol. XXIII, p. 67.

has the value 1: but the double integral does not exist, since $I(K) > 0$, for any value of k in the interval (0, 1).

2.* Let x be represented by a finite or infinite decimal, excluding those decimals in which every figure from and after some fixed place is 9. Let p_x denote the number of decimal places in the representation of x in the manner described. Let y be represented in a similar manner, with a corresponding definition of p_y. Let the function $f(x, y)$ be defined in the rectangle bounded by $x=0$, $x=1$, $y=0$, $y=1$, by $f(x, y)=\frac{1}{p_x+1}+\frac{1}{p_y+1}$ when p_x and p_y are both finite; otherwise let $f(x, y)=0$.

We have $\int_0^1 \frac{1}{p_y+1}\,dy=0$; for there are only a finite number of values of y in (0, 1), for which p_y is less than an arbitrarily chosen fixed integer, or $\frac{1}{p_y+1}$ is greater than an arbitrarily chosen fixed proper fraction. The function $f(x, y)$ vanishes, except when one at least of x and y is representable by a finite decimal; and thus the double integral

$$\int f(x, y)\,(dx\,dy)=0.$$

Now $$\overline{\int_0^1} f(x, y)\,dy=\frac{1}{p_x+1}, \qquad \underline{\int_0^1} f(x, y)\,dy=0;$$

and thus $\int_0^1 f(x, y)\,dy$ has no definite value for any value x of the everywhere-dense enumerable set of points for which p_x is finite.

Nevertheless $$\int_0^1 dx \overline{\underline{\int_0^1}} f(x, y)\,dy=0=\int f(x, y)\,(dx\,dy).$$

3*. With the same notation as in the last example, let $f(x, y)=0$, when p_x, p_y are both finite or both infinite; let $f(x, y)=\frac{1}{1+p_x}$, when p_x is finite and p_y infinite, and $f(x, y)=\frac{1}{1+p_y}$, when p_y is finite and p_x is infinite. In this case $f(x, y)$ differs from 0 at an unenumerable set of points; and yet the set of points at which $f(x, y)>\epsilon$ has the plane content zero, since all such points are on a finite number of lines parallel to the coordinate axis, although they are everywhere-dense on those lines. The double integral, and consequently the repeated integrals, exist in this case.

4.* An example has been given in Ex. 1, § 108, of a set of points K which is everywhere-dense and unclosed, whereas the sets K_x, K_y are all finite. Let $f(x, y)=c'$ at every point of K, and $=c$ at every other point. In this case, the double integral does not exist; but

$$\int_0^1 f(x, y)\,dx=c, \qquad \int_0^1 f(x, y)\,dy=c,$$

whatever values y and x may have in the first and in the second integral respectively.

In this case $$\int_0^1 dx \int_0^1 f(x, y)\,dy \text{ and } \int_0^1 dy \int_0^1 f(x, y)\,dx$$

both exist and have the same value c.

5.* Let a set $\{(x', y')\}$ be defined as follows:—Let x' have any value for which $p_{x'}$ is finite; and with such a fixed value of x', let every y' be taken for which $p_{y'} \leq p_{x'}$. On every line parallel to the y-axis there are only a finite number of points of the set; but

* Pringsheim, *Sitzungsberichte d. Leipziger Akademie*, vol. XXVIII, p. 71.

the set is everywhere-dense on every line which is parallel to the x-axis, and which has for its ordinate one of the y'. Let $f(x, y)=c'$ for the set $\{(x', y')\}$, and let $f(x, y)=c$ for all remaining points.

We have, in this case,
$$\int_0^1 f(x, y)\,dy=c,$$
and
$$\int_0^1 dx \int_0^1 f(x, y)\,dy=c.$$

But $\int_0^1 f(x, y')\,dx$ has c and c' for its upper and lower values; and the set of values of y' being everywhere-dense, $\int_0^1 dy \int_0^1 f(x, y)\,dx$ does not exist.

6.* Let $f(x, y)$ be defined at all points of the rectangle bounded by $x=0$, $x=1$, $y=0$, $y=1$, by the condition that $f(x, y)=0$, except at those points (x', y') at which

$$x'=\frac{2m+1}{2^n}, \quad y'=\frac{2p+1}{2^q}, \text{ where } f(x', y')=\frac{1}{2^n},$$

m, n, p and q being positive integers.

In this case the double integral exists, and therefore the repeated integrals both exist†.

PROPERTIES OF THE DOUBLE INTEGRAL.

317. The following properties of the proper double integral may be established in a simple manner:—

(1) If $f(x, y)$ have a double integral in the domain G, then $|f(x, y)|$ also has a double integral in the same domain; *i.e.* $f(x, y)$ is absolutely integrable in the domain G.

For the fluctuation of $|f(x, y)|$ in any rectangular cell cannot exceed that of $f(x, y)$; hence it follows from the condition of § 311, that $|f(x, y)|$ is integrable, if $f(x, y)$ be so.

It follows from the definition that

$$\left|\int_G f(x,y)(dxdy)\right| \leqq \int_G |f(x, y)|\,(dxdy).$$

(2) If G be divided into two parts G_1, G_2, each of which is measurable in accordance with Jordan's definition, then if $f(x, y)$ have a double integral in G, it has also double integrals in G_1 and G_2, which satisfy the condition

$$\int_{G_1} f(x, y)\,(dxdy) + \int_{G_2} f(x, y)\,(dxdy) = \int_G f(x, y)\,(dxdy).$$

For K, the set of points of G at which the saltus of $f(x, y)$ is $\geqq k$, may be divided into its two components K_1 in G_1, and K_2 in G_2. The only points of

* Du Bois Reymond, *Crelle's Journal*, vol. XCIV, p. 278; also Stolz, *Grundzüge*, vol. III, p. 73.

† This is denied by Stolz, *Grundzüge*, vol. III, p. 88, on the ground that $f(x, y)$ for $x=\frac{2m+1}{2^n}$ is not integrable with respect to y, but has $\frac{1}{2^n}$ and 0 for its upper and lower values. We have, however, shown that this is no justification for denying the existence of the repeated integral.

G_1 at which the saltus of $f(x, y)$, considered as defined for G_1 only, is $\geqq k$, consist of the points of the set K_1, together with points forming a set K_1' on the frontier of G_2. The content of this frontier being zero, K_1' has zero content; also K_1 has zero content, since it is a part of K. Since $K_1 + K_1'$ has, for every value of k, the content zero, it follows that $f(x, y)$ is integrable in G_1; similarly, it is integrable in G_2. Also

$$\int_G f(x, y)\,(dxdy)$$

is, by definition, the limit of the finite sum

$$\delta_1 M(\delta_1) + \delta_2 M(\delta_2) + \ldots + \delta_n M(\delta_n).$$

The rectangles δ consist of (1) those which contain interior points of G_1 only, (2) those which contain interior points of G_2 only, (3) those which contain points on the frontiers of G and of G_1 and G_2. The above sum may therefore be divided into three portions containing those δ's respectively which belong to (1), (2) and (3). The limit of the first of these is

$$\int_{G_1} f(x, y)\,(dxdy),$$

that of the second is

$$\int_{G_2} f(x, y)\,(dxdy),$$

and that of the third is at most equal to U multiplied by the content of the points on the frontiers of G_1 and G_2, where U is the upper limit of $|f(x, y)|$. Since the contents of the frontiers are zero, the limit of the third part of the sum is zero; hence the second part of the theorem is established.

(3) If $F(f_1, f_2, \ldots f_n)$ be a continuous function in G of the n functions $f_1, f_2, \ldots f_n$, each of which has a double integral in G, then F has itself a double integral in G.

The proof of this is identical with the one applicable to the case of a single variable, given in § 256.

That the sum or product of two or more integrable functions is integrable is a particular case of this theorem.

(4) If f be integrable in G, and the function f_1 be defined by $f_1 = f$, for every positive value of f, and by $f_1 = 0$, when f is negative or zero; the function f_2 being defined by $f_2 = -f$, for every negative value of f, and by $f_2 = 0$, when f is zero or positive; then f_1 and f_2 are each integrable in G. For $f_1 - f_2$ being $= f$, is integrable, and also in virtue of (1), $f_1 + f_2$ is integrable in G, hence by (3), f_1 and f_2 are both integrable in G.

IMPROPER DOUBLE INTEGRALS.

318. As in the case of single integrals, the definition of a double integral may be extended to the case in which the function has a set of points of infinite discontinuity. This set is necessarily closed, and it will be assumed throughout that its plane content is zero. It will also be assumed that the domain for which such a function is defined is bounded, and that the frontier has the content zero, the domain being therefore measurable in accordance with Jordan's definition; and consequently the function may be replaced by another function defined for all the points in a fundamental rectangle, the new function being taken to vanish at all points not in the original domain, and to have the same value as the original function at all points of that domain. These assumptions being made, a definition of the improper double integral which is substantially the one given by Jordan*, and adopted by Stolz†, may be stated as follows:—

Let $D_1, D_2, \ldots D_n \ldots$ denote a sequence of domains contained in the fundamental rectangle, each one of which consists of a finite number of connex closed portions each with its frontier of zero content, and in which the number of the portions may increase indefinitely with n. Further, let us suppose that none of these domains contain, in their interiors or on their frontiers, any point at which $f(x, y)$ has an infinite discontinuity, and that the sequence is such that the measure of D_n converges to that of the fundamental rectangle; then if the upper integrals

$$\overline{\int}_{D_1} f(x, y)(dxdy),\ \overline{\int}_{D_2} f(x, y)(dxdy), \ldots \overline{\int}_{D_n} f(x, y)(dxdy), \ldots$$

taken over the domains $D_1, D_2, \ldots$, converge to a definite limit independent of the particular sequence $\{D_n\}$ chosen, this limit is defined to be the improper upper integral

$$\overline{\int} f(x, y)(dxdy)$$

of $f(x, y)$ in the given domain. A similar statement applies to the case of the improper lower integral. When the improper upper and lower integrals both exist, and have the same value, then the improper integral

$$\int f(x, y)(dxdy)$$

over the given domain is said to exist, and to have this common value.

It will be observed that the domains D_n are all measurable in accordance with Jordan's definition of a measurable set, and therefore also in accordance with the definition of Borel and Lebesgue.

* *Cours d'Analyse*, vol. II, p. 76.

† *Grundzüge*, vol. III, p. 124.

In case the function $f(x, y)$ be integrable in all the domains $D_1, D_2, \ldots$, however this sequence may be chosen, subject to the conditions stated above, then if the sequence

$$\int_{D_1} f(x, y)\,(dx\,dy),\ \int_{D_2} f(x, y)\,(dx\,dy), \ldots$$

converge to a definite limit, independent of the particular sequence $\{D_n\}$, that limit defines the improper double integral

$$\int f(x, y)\,(dx\,dy).$$

If a function $f(x, y)$ have an improper integral in the fundamental rectangle, then $f(x, y)$ has a proper integral in any connex closed domain of which the frontier has zero measure, and which is contained in the fundamental rectangle, but itself contains no points in its interior, or on its frontier, at which $f(x, y)$ is infinitely discontinuous. For, by the definition,

$$\overline{\int}_D f(x, y)\,(dx\,dy),\ \underline{\int}_D f(x, y)\,(dx\,dy)$$

converge to one and the same definite limit, as D converges to the fundamental rectangle; therefore D can be so chosen that, if ϵ be an arbitrarily chosen positive number,

$$\overline{\int}_D f(x, y)\,(dx\,dy) - \underline{\int}_D f(x, y)\,(dx\,dy) < \epsilon.$$

Now if D' be any domain of the type defined above, in the interior of D, it is clear that the difference between the upper and lower integrals of $f(x, y)$ throughout D' cannot exceed the difference of the upper and lower integrals throughout D, and is therefore $< \epsilon$. Since ϵ is arbitrarily small, it follows that the upper and lower integrals throughout D' must be identical, and therefore that $f(x, y)$ is integrable in D'. It has thus been shewn that *if $f(x, y)$ have an improper double integral in the fundamental rectangle, it must possess a proper integral in any connex closed domain interior to that rectangle, such that the domain has its frontier of zero measure, and contains no points of infinite discontinuity of the function, either in its interior or on its frontier.*

319. *The necessary and sufficient condition for the existence of the improper upper double integral*

$$\overline{\int} f(x, y)\,(dx\,dy)$$

is that, corresponding to any arbitrarily chosen positive number ϵ, another positive number δ can be determined, such that, if Δ be any connex closed domain whatever, of which the frontier has zero measure, and which is contained *in the fundamental rectangle, but itself contains no points of infinite discontinuity of* $f(x, y)$, *either in its interior or on its frontier, then, provided the measure of* Δ *is* $< \delta$, *the condition*

$$\left|\overline{\int} f(x, y)(dxdy)\right| < \epsilon,$$

taken over Δ, *is satisfied.*

A similar theorem applies to the improper lower double integral.

To shew that the condition stated in the theorem is sufficient, let D, D' be two domains of the kind specified in § 318, such that $m(D)$, $m(D')$ both differ from the area of the fundamental rectangle by less than δ; they are both interior to the fundamental rectangle, and contain none of the points of infinite discontinuity of the function. Let d be the set of points of D which do not belong to D', and d' the set of points of D' which do not belong to D; then

$$m(d) < A - m(D') < \delta,$$

and
$$m(d') < A - m(D) < \delta,$$

where A is the area of the fundamental rectangle. Also, since the domains $D + d'$, $D' + d$ are identical, we have

$$I_D - I_{D'} = I_d - I_{d'},$$

where I denotes the upper double integral

$$\overline{\int} f(x, y)(dxdy)$$

taken over the domain indicated by a suffix. It follows that

$$|I_D - I_{D'}| \leqq |I_d| + |I_{d'}| < 2\epsilon;$$

and hence it is easily seen that any two sequences

$$\{I_{D_n}\}, \quad \{I_{D'_n}\}$$

both converge to one and the same definite limit.

To shew that the condition stated in the theorem is necessary, let us suppose that it is not satisfied. We thus assume that a domain d of arbitrarily small measure can be found, such that $|I_d| > \epsilon$.

Let D be interior to the rectangle, and such that

$$A - m(D) < \delta.$$

Taking D to contain d, we then have

$$m(D - d) > A - 2\delta,$$

provided
$$m(d) < \delta.$$

The two domains D, $D - d$ both converge to A, if δ be decreased indefinitely; and
$$I_D - I_{D-d} = I_d;$$

thus $$| I_D - I_{D-d} | > \epsilon,$$
however small δ may be; hence the limit does not in this case exist.

The necessary and sufficient condition that the improper upper and lower integrals of $f(x, y)$ in the fundamental rectangle may both exist is that the improper upper integral of $|f(x, y)|$ may exist.

To shew that the condition stated is sufficient, we observe that, on the assumption of the existence of
$$\overline{\int} | f(x, y) | (dx\,dy),$$
it follows from the theorem established above that the upper integral of
$$| f(x, y) |$$
through a connex domain D, interior to A, and containing none of the points of infinite discontinuity, tends to the limit zero as $m(D)$ does so. Also
$$\left| \overline{\int}_D f(x, y)\,(dx\,dy) \right|, \quad \left| \underline{\int}_D f(x, y)\,(dx\,dy) \right|$$
are both $$\leqq \overline{\int}_D | f(x, y) | (dx\,dy),$$
as is easily seen. It follows that both
$$\overline{\int}_D f(x, y)\,(dx\,dy), \quad \underline{\int}_D f(x, y)\,(dx\,dy)$$
converge to zero as $m(D)$ does so, and uniformly for all such domains D; and that these are the sufficient conditions for the existence of
$$\overline{\int} f(x, y)\,(dx\,dy), \quad \underline{\int} f(x, y)\,(dx\,dy).$$

To prove that the condition stated is a necessary one, let us assume that, for every connex domain D satisfying the specified conditions, and such that $m(D) < \delta$, we have
$$\left| \overline{\int}_D f(x, y)(dx\,dy) \right| < \epsilon.$$
Now let $$f(x, y) = f^+(x, y) - f^-(x, y),$$
where $$f^+(x, y) = f(x, y)$$
at all points where $f(x, y)$ is positive, and everywhere else
$$f^+(x, y) = 0;$$
also $$f^-(x, y) = -f(x, y),$$
at every point where $f(x, y)$ is negative, and $f^-(x, y)$ is everywhere else zero. The domain D may be divided into a finite number of rectangles δ, some of which may lie partly outside D; the functions being taken to be

zero in all such outlying portions. Denoting by U_δ the upper limit of a function in the rectangle δ, we have

$$U_\delta\{f(x, y)\} = U_\delta\{f^+(x, y)\},$$

in all elements δ in which $f(x, y)$ has positive values; and in all other elements

$$U_\delta\{f^+(x, y)\} = 0.$$

The elements may be taken such that, if η be an arbitrarily chosen number, the inequalities

$$\Sigma\delta U_\delta\{f(x, y)\} - \overline{\int_D} f(x, y)(dxdy) < \eta,$$

$$\Sigma\delta U_\delta\{f^+(x, y)\} - \overline{\int_D} f^+(x, y)(dxdy) < \eta$$

are both satisfied. These conditions are also satisfied for any domain contained in D.

We have now

$$\overline{\int_D} f^+(x, y)\,dxdy \leqq \sum_D \delta U_\delta\{f^+(x, y)\} \leqq \sum_{D_1} \delta U_\delta\{f(x, y)\}$$
$$\leqq \overline{\int_{D_1}} f(x, y)(dxdy) + \eta$$
$$\leqq \epsilon + \eta,$$

where D_1 consists of the domain which is composed of those elements δ of D in which $f(x, y)$ has positive values. Since η is arbitrarily small, we have

$$\overline{\int_D} f^+(x, y)(dxdy) \leqq \epsilon.$$

Again, since

$$\underline{\int_D} f(x, y)(dxdy) = -\overline{\int_D} - \{f(x, y)\}(dxdy)$$

we see that

$$\overline{\int_D} - \{f(x, y)\}(dxdy)$$

has the limit zero when $m(D)$ has the limit zero; and from this it follows as before that

$$\overline{\int_D} f^-(x, y)(dxdy) \leqq \epsilon.$$

Since

$$f(x, y) = f^+(x, y) - f^-(x, y),$$

we have

$$\overline{\int_D} |f(x, y)|(dxdy) \leqq 2\epsilon;$$

and therefore

$$\overline{\int_D} |f(x, y)|(dxdy)$$

has the limit zero when $m(D)$ converges to zero. It now appears, by employing the first theorem of this section, that

$$\overline{\int}|f(x, y)|\,(dxdy),$$

taken throughout the fundamental rectangle, has a definite value.

It should be observed that since

$$\underline{\int}_D |f(x, y)|\,(dxdy) \leqq \overline{\int}_D |f(x, y)|\,(dxdy),$$

the lower integral

$$\underline{\int}_D |f(x, y)|\,(dxdy)$$

has the limit zero, whenever zero is the limit of

$$\overline{\int}_D |f(x, y)|\,(dxdy);$$

and that therefore $\underline{\int}|f(x, y)|\,(dxdy)$

always exists when $\overline{\int}|f(x, y)|\,(dxdy)$

does so, the integration being over the fundamental rectangle.

320. It has been seen in § 318 that, in case $f(x,y)$ have an improper integral in the fundamental rectangle, it has a proper integral in any closed connex domain D contained in that rectangle, with frontier of zero measure, and containing no points of infinite discontinuity of the function. It follows by the theorem of § 317 (1), that $|f(x,y)|$ is also integrable in the domain D; and we have already seen that the existence of an improper upper integral of $|f(x, y)|$ is a necessary consequence of the existence of the improper integral of $f(x, y)$. It thus appears that the improper upper, and lower, integrals of $|f(x, y)|$ must be identical, and therefore that, *if $f(x, y)$ be a function which has an improper double integral, in accordance with Jordan's definition, then $|f(x, y)|$ has also an improper integral, so that every such improper integral is absolutely convergent.*

We have seen that Harnack's definition of an improper single integral is applicable not only to the cases in which the convergence is absolute, but also to cases in which the convergence is not absolute. Jordan's definition of an improper double integral is however much more stringent than Harnack's definition of an improper single integral. In the latter case the integral is defined as the limit of the proper integral taken through a finite number of intervals, not chosen arbitrarily in any manner consistent with the condition that the sum of these intervals is to converge to the length of the interval of integration, but chosen so as to satisfy the special condition that they are complementary to a finite set of intervals which contain in their interiors all the points of infinite discontinuity of the function, each interval of the finite set containing at least one such point.

If the proper integrals of which the improper integral, in Harnack's definition, is the limit, were not subjected to the above mentioned restriction, reasoning precisely similar to that applied above would shew that every improper single integral must be absolutely convergent. In order that a definition of the improper double integral should admit of the existence of double integrals which do not converge absolutely, it would be necessary to subject the domains $D_1, D_2, \dots D_n, \dots$, (the proper integrals through which define, as their limit, the double integral), to some restriction which would allow of the existence of a limit in cases in which such a limit does not exist, independently of the particular set $\{D_n\}$ chosen, when no such restriction is made. Such a restriction as to the nature of the domains D_n would correspond to the restriction to a special class of sets of intervals, of the intervals through which the proper integrals in Harnack's definition of an improper single integral are taken. The true extension of Harnack's definition to the case of double integrals would be the following:—

Let the points of infinite discontinuity of $f(x, y)$ (the set of such points being of zero content), be enclosed in a finite set of rectangles with sides parallel to those of the fundamental rectangle, each rectangle of the finite set containing at least one point of infinite discontinuity, and no such point being on the frontier of the set of rectangles; and let D_n denote the remaining part of the fundamental rectangle when the finite set of rectangles is removed, then, if $f(x, y)$ have a proper integral in every such domain D_n, and if this proper integral converge to a definite limit when any sequence whatever of such domains D_n is taken, such that the measure of D_n converges to that of the fundamental rectangle, this limit shall define the improper double integral of $f(x, y)$.

This extension of Harnack's definition would admit of the existence of non-absolutely convergent improper double integrals, as we have seen to be the case for improper single integrals. With this definition, the theorems of § 319 would no longer be valid.

When it is asserted that non-absolutely convergent double integrals do not exist, the assertion must be taken to mean that such integrals do not exist in accordance with the definition of Jordan, and not that it is impossible to give definitions, such as the above extension of Harnack's, in accordance with which double integrals exist that do not converge absolutely.

The properties of improper double integrals which are not necessarily absolutely convergent are more restricted than those which exist in accordance with Jordan's definition, and it is consequently a matter of opinion whether, though the former certainly exist as limits, the name integral may be appropriately applied to such limits.

EXAMPLE.

If we take as the domain of integration the rectangle bounded by $x=0$, $x=a$, $y=0$, $y=b$, then the double integral of $\frac{1}{x}\sin\frac{1}{x}$ is not convergent, and therefore in accordance with Jordan's definition does not exist; although the single integral $\int_0^a \frac{1}{x}\sin\frac{1}{x}\,dx$ is non-absolutely convergent, and does exist in accordance with Harnack's definition.

The existence of $\int_0^a \frac{1}{x}\sin\frac{1}{x}\,dx$ depends upon the fact that $\int_\epsilon^a \frac{1}{x}\sin\frac{1}{x}\,dx$ converges to a definite limit as ϵ converges to zero, and this is sufficient to ensure the existence of the single integral. Although $\int \frac{1}{x}\sin\frac{1}{x}\,(dx\,dy)$ taken over the domain bounded by $x=\epsilon$, $x=a$, $y=0$, $y=b$, converges to a definite limit, as ϵ converges to zero, this is not sufficient to ensure the existence of the Jordan double integral. Taking Jordan's definition, let the domain D_n consist of the rectangular spaces bounded by the lines $y=0$, $y=b$, and the lines parallel to the y-axis at the extremities of the intervals on the x-axis

$$\left(\frac{1}{(2n+1)\pi}, a\right), \quad \left(\frac{1}{(2n+3)\pi}, \frac{1}{(2n+2)\pi}\right), \quad \left(\frac{1}{(2n+5)\pi}, \frac{1}{(2n+4)\pi}\right), \ldots \left(\frac{1}{(4n+1)\pi}, \frac{1}{4n\pi}\right).$$

The double integral taken through these spaces is

$$\int_{\frac{1}{(2n+1)\pi}}^{a} \int_0^b \frac{1}{x}\sin\frac{1}{x}\,dx\,dy + \sum_{p=n+1}^{p=2n} \int_{\frac{1}{(2p+1)\pi}}^{\frac{1}{2p\pi}} \int_0^b \frac{1}{x}\sin\frac{1}{x}\,dx\,dy,$$

or

$$\int_{\frac{1}{(2n+1)\pi}}^{a} \int_0^b \frac{1}{x}\sin\frac{1}{x}\,dx\,dy + b\int_0^\pi \sin z \sum_{p=n+1}^{p=2n} \frac{1}{z+2p\pi}\,dz;$$

which is greater than

$$b\int_{\frac{1}{(2n+1)\pi}}^{a} \frac{1}{x}\sin\frac{1}{x}\,dx + b\frac{n}{(4n+1)\pi}\int_0^\pi \sin z\,dz,$$

or than

$$b\int_{\frac{1}{(2n+1)\pi}}^{a} \frac{1}{x}\sin\frac{1}{x}\,dx + \frac{2nb}{(4n+1)\pi};$$

and this converges to

$$b\int_0^a \frac{1}{x}\sin\frac{1}{x}\,dx + \frac{b}{2\pi}:$$

whereas $\int \frac{1}{x}\sin\frac{1}{x}\,(dx\,dy)$ taken over the domain bounded by $x=\epsilon$, $x=a$, $y=0$, $y=b$, converges to

$$b\int_0^a \frac{1}{x}\sin\frac{1}{x}\,dx.$$

Therefore the mode of choice of the intervals D_n affects the limit to which

$$\int_{D_n} \frac{1}{x}\sin\frac{1}{x}\,(dx\,dy)$$

converges, as D_n converges to the complete domain. Thus it is clear that the double integral, in accordance with Jordan's definition, does not exist.

321. A definition of the improper double integral has been given by de la Vallée-Poussin, precisely similar to his definition for single integrals. This may be stated as follows:—

Let $f_n(x, y)$ *be the function which is such that*

$$f_n(x, y) = f(x, y)$$

at every point (x, y) *at which*

$$M_n \geqq f(x, y) \geqq -N_n,$$

where M_n, N_n *are two positive numbers, and is also such that*

$$f_n(x, y) = M_n,$$

at every point where $f(x, y) > M_n$; *also let* $f_n(x, y) = -N_n$,

at every point where $\qquad f(x, y) < -N_n.$

If $f(x, y)$ *be such that the proper integral*

$$\int f_n(x, y)\,(dxdy)$$

exists, whatever positive values M_n, N_n *may have, and if also the double limit*

$$\lim_{M_n=\infty,\ N_n=\infty} \int f_n(x, y)\,(dxdy)$$

have a definite finite value, that limit is said to define the improper integral

$$\int f(x, y)\,(dxdy).$$

The existence of the double limit implies that, if

$$M_1, M_2, \dots M_n, \dots \text{ and } N_1, N_2, \dots N_n, \dots$$

be any two independent sequences of increasing numbers with no upper limits, then the sequence of numbers

$$\int f_1(x, y)\,(dxdy),\quad \int f_2(x, y)\,(dxdy) \dots \int f_n(x, y)\,(dxdy) \dots$$

converges to a definite limit, independently of the mode in which the two sequences $\{M_n\}$, $\{N_n\}$ are chosen.

It may be shewn, exactly in the same manner as in § 273, that the existence, in accordance with this definition, of the improper integral

$$\int f(x, y)\,(dxdy),$$

implies the existence of $\qquad \int |f(x, y)|\,(dxdy)$;

and thus, that all improper integrals, so defined, are absolutely convergent.

As in § 273, it appears that, for the existence of the double integral, it is necessary that the closed set of points of infinite discontinuity of the function should have zero content; and in fact that, k denoting any positive number, the set of points at which the measure of discontinuity of the function is $\geqq k$, must have content zero. Thus all the points of discontinuity of the function form a set of zero plane measure.

322. Since the definitions given by de la Vallée-Poussin and by Jordan both apply only to the case of absolutely convergent integrals, it is of importance to shew that they are completely equivalent to one another. If, in the proof given in § 274, of the equivalence of the two definitions of absolutely convergent single integrals, intervals be replaced throughout by rectangular cells with sides parallel to the sides of the fundamental rectangle, we have a proof that in the case of absolutely convergent double integrals, the definition of de la Vallée-Poussin is completely equivalent to the extended definition of Harnack, given in § 320. It only remains to prove that the latter is, in the present case, equivalent to Jordan's. It will be sufficient to give a proof that, for a function which is never negative, if the integrals taken through the special domains D_n, which consist of sets of rectangles with sides parallel to the axes of x and y, converge to a definite limit, then the integrals taken through domains D'_n, each of which consists of a finite number of connex closed portions of any kind, also converge to the same definite value as in the case of D_n, when $m(D'_n)$ converges to the area A of the fundamental rectangle.

Taking D'_n so that $A - m(D'_n) = \epsilon_n$, D_n can be so chosen as to contain D'_n in its interior. For, since D'_n does not contain either in its interior or on its frontier any points of infinite discontinuity of the function, therefore, for each one of the latter points, the distance from all the points of D'_n has a minimum greater than zero, D'_n being closed and connex. Hence each point of infinite discontinuity can be enclosed in a rectangle which contains no points of D'_n either in its interior or on its sides.

Since the set of points of infinite discontinuity is closed, a finite set of the rectangles can, in accordance with the Heine-Borel theorem, be chosen so as to enclose the whole set of these points; and the complement of this finite set of rectangles may be taken to be D_n.

This may be done for each value of n.

If $m(D'_n)$ converge to A, it is clear that $m(D_n)$ which is $> m(D'_n)$ also converges to A. Also a number $n' > n$, can be determined, such that $D'_{n'}$ encloses D_n in its interior; we have then

$$\int_{D'_{n'}} f(x, y)(dxdy) \geqq \int_{D_n} f(x, y)(dxdy) \geqq \int_{D'_n} f(x, y)(dxdy).$$

If
$$\int_{D_n} f(x, y)(dxdy)$$

converge to a definite limit $\int f(x, y)(dxdy)$,

n may be taken so great that

$$\int f(x, y)(dxdy) - \int_{D_n} f(x, y)(dxdy)$$

is less than the arbitrarily small number η; then also

$$\int f(x, y)(dxdy) - \int_{D'_{n'}} f(x, y)(dxdy) < \eta,$$

and it thus follows that $\int_{D'_n} f(x, y)(dxdy)$

also converges to the limit $\int f(x, y)(dxdy)$.

The complete equivalence of the two definitions has now been established.

The definition of a Lebesgue integral may be extended to the case of improper integrals, as in § 291. It may be shewn, precisely as in § 291, that the definition of a Lebesgue improper double integral may be put into a form which is an extension of de la Vallée-Poussin's definition to the case in which the functions $f_n(x, y)$ possess Lebesgue integrals, but not necessarily Riemann integrals. Exactly as in § 291, it may be shewn that the three definitions of an improper double integral are in accord* with one another whenever all three are applicable.

DOUBLE INTEGRALS OVER INFINITE DOMAINS.

323. Let the function $f(x, y)$ be defined for an unbounded domain G, and let it be assumed that $f(x, y)$ possesses either proper, or improper, upper and lower integrals for every bounded domain D contained in G, such domain being closed, and having its frontier of zero content.

Let† $D_1, D_2, \ldots D_n, \ldots$ *be a sequence of domains each consisting of a finite number of connex closed portions, and such that* D_n *contains every point of* G *of which the distance from the origin is* $< \rho_n$, *where* ρ_n *is a positive number which increases indefinitely with* n. *If the proper or improper upper integral of* $f(x, y)$ *taken over* D_n *have, for the whole sequence, a definite limit independent of the particular sequence* $\{D_n\}$ *chosen, subject to the above condition, then this limit is said to define the improper upper integral of* $f(x, y)$ *over the unbounded domain* G, *and it may be denoted by* $\overline{\int}_G f(x, y)(dxdy)$. *A similar statement applies to the improper lower integral* $\underline{\int}_G f(x, y)(dxdy)$.

* Hobson, *Proc. Lond. Math. Soc.* ser. 2, vol. IV, p. 145.

† See Jordan's *Cours d'Analyse*, vol. II, p. 81; also Stolz's *Grundzüge*, vol. II, p. 148.

When the improper upper and lower integrals both exist, and have the same value, then this value defines the improper integral

$$\int_G f(x, y)\,(dx\,dy)$$

of $f(x, y)$ over the domain G.

The necessary and sufficient condition for the existence of

$$\overline{\int}_G f(x, y)\,(dx\,dy)$$

is that, corresponding to each arbitrarily chosen number ϵ, a number ρ can be determined, such that for every bounded connex closed domain Δ contained in G, and itself containing no points of distance from the origin $< \rho$, the inequality

$$\left|\overline{\int}_\Delta f(x, y)\,(dx\,dy)\right| < \epsilon$$

be satisfied.

Let us first assume that the condition stated is fulfilled. Let D, D' be two domains each containing as a part all those points of G of which the distance from the origin is $< \rho$, and let E be the domain common to D and D'.

Let $D - E = \Delta$, $D' - E = \Delta'$; then all points of Δ, Δ' are at distances $\geqq \rho$ from the origin. We have then

$$\left|\overline{\int}_\Delta f(x, y)\,(dx\,dy)\right| < \epsilon, \qquad \left|\overline{\int}_{\Delta'} f(x, y)\,(dx\,dy)\right| < \epsilon;$$

and since $\Delta - \Delta' = D - D'$, we have

$$\left|\overline{\int}_D f(x, y)\,(dx\,dy) - \overline{\int}_{D'} f(x, y)\,(dx\,dy)\right| < 2\epsilon;$$

and the condition for the existence of

$$\overline{\int}_G f(x, y)\,(dx\,dy)$$

is therefore satisfied.

To shew that the condition stated in the theorem is a necessary one, let us assume that the condition is not satisfied. Then, for every pair of values of ϵ and ρ, there exists a domain Δ, all the points of which are of distance $\geqq \rho$ from the origin, such that

$$\left|\overline{\int}_\Delta f(x, y)\,(dx\,dy)\right| \geqq \epsilon.$$

Let E be a bounded domain contained in G, and itself containing every point of which the distance from the origin is $< \rho$; and let E contain Δ. Let

$E - \Delta = E_1$; then E_1 also contains every point of G whose distance from the origin is $< \rho$. Then

$$\left| \overline{\int}_E f(x, y)\,(dx\,dy) - \overline{\int}_{E_1} f(x, y)\,(dx\,dy) \right| \geqq \epsilon.$$

It follows that, however great ρ may be, there exist in G two domains, each containing all points of G of distance from the origin $< \rho$, such that for them this last condition is satisfied; in this case the improper upper integral through G cannot exist. It has therefore been shewn that the condition stated in the theorem is a necessary one.

The necessary and sufficient condition that the upper and lower integrals of $f(x, y)$ through the infinite domain G may both exist is that the improper upper integral of $|f(x, y)|$ over G may exist.

The proof of this theorem is a repetition of the proof of the corresponding theorem in § 319; the only difference being that, in the present case, D is taken to be a connex closed domain contained in G, every point of which is at a distance from the origin $\geqq \rho$. The indefinite increase of ρ corresponds to the indefinite diminution of $m(D)$ in the former case.

324. It now follows, as in § 320, that, *if $f(x, y)$ be a function which has an improper integral over the infinite domain G, in accordance with Jordan's definition, then $|f(x, y)|$ has also such an improper integral, so that every improper integral is absolutely convergent.*

In order to obtain a definition in accordance with which non-absolutely convergent improper integrals over an infinite domain may exist, some restriction must be imposed upon the nature of the domains $D_1, D_2, \ldots$ which are employed in the definition in § 323. For example, we may restrict $D_1, D_2, \ldots$, as in the extended definition of Harnack, which has been given in § 320.

EXAMPLES.

1*. The integral

$$\int \sin(ax + by)\, x^{r-1} y^{s-1}\,(dx\,dy),$$

where $0 < r < 1$, $0 < s < 1$, taken over the positive quadrant, has no existence as an absolutely convergent improper integral. We find that the integral taken over the rectangle bounded by $x = 0$, $x = h$, $y = 0$, $y = k$ tends to the limit $a^{-r} b^{-s} \Gamma(r) \Gamma(s) \sin \frac{1}{2}(r + s)\pi$, as h and k are increased indefinitely. If the integral be taken over the domain $x > 0$, $y > 0$, $ax + by < h$, then when h is indefinitely increased, the integral has no limit if $1 < r + s < 2$; but it tends to the same limit as before, when $r + s < 1$. The integral may be regarded as conditionally convergent, if we adopt a definition in accordance with which it is sufficient that the integral taken through the rectangle $x = 0$, $x = h$, $y = 0$, $y = k$ have a definite double limit, as h and k are indefinitely increased.

* Hardy, *Messenger of Math.*, vol. XXXII, p. 96.

2*. The integrals

$$\int \cos(ax^2+2hxy+by^2)\,(dxdy), \qquad \int \sin(ax^2+2hxy+by^2)\,(dxdy)$$

where a, $ab-h^2$ are positive, taken over the positive quadrant, do not exist as absolutely convergent integrals. It may be seen that, if the integrals are taken over the quadrant of a circle bounded by $r=R$, the value of the integral has no definite limit as R is increased indefinitely. If the integral be taken over the rectangle bounded by $x=0$, $x=h'$, $y=0$, $y=k'$, then, when h' and k' are increased indefinitely, the integrals have

$$0, \text{ and } \frac{1}{2\sqrt{ab-h^2}}\cos^{-1}\frac{h}{\sqrt{ab}}$$

for limits respectively, the inverse cosine having its least positive value. These may be regarded as the values of the integrals, subject to a suitable restriction on the domains of which the positive quadrant is the limit.

If $a=0$, $b=0$, $h=\frac{1}{2}$,

$$\int \sin xy\,(dxdy)$$

over the positive quadrant, has no existence, even considered as the limit of an integral over the rectangle. But

$$\int \cos xy\,(dxdy)$$

exists and is equal to $\frac{1}{2}\pi$, when the integral is defined as the limit of the integral over the finite rectangle. It may be remarked that the single integrals

$$\int_0^{h'} \cos xy\,dx, \qquad \int_0^{k'} \sin xy\,dy$$

are both divergent.

THE TRANSFORMATION OF DOUBLE INTEGRALS.

325. Let (x, y) be a point of a limited perfect and connex domain H, and let x and y be expressed by means of two functions f_1, f_2 in terms of two new variables ξ, η, which may be represented by points (ξ, η) in another plane. Let us suppose that the functions

$$x=f_1(\xi, \eta), \qquad y=f_2(\xi, \eta),$$

and the reciprocal functions

$$\xi=\phi_1(x, y), \qquad \eta=\phi_2(x, y),$$

are such that the following conditions are satisfied:—

(1) To each point (x, y) there corresponds one point (ξ, η); and, conversely, to each point (ξ, η) there corresponds one point (x, y); and to the limited domain H there corresponds a limited domain $\overline{H}$.

(2) The functions $f_1(\xi, \eta)$, $f_2(\xi, \eta)$ are continuous functions of (ξ, η) throughout the domain $\overline{H}$.

* Hardy, *Messenger of Math.*, vol. XXXII, p. 159.

(3) The functions $f_1(\xi, \eta)$, $f_2(\xi, \eta)$ have, at every point (ξ, η) of $\bar{H}$, definite partial differential coefficients with respect to ξ and η, and each one of these is everywhere continuous with respect to (ξ, η), and nowhere vanishes.

(4) The Jacobian of $f_1(\xi, \eta)$, $f_2(\xi, \eta)$ with respect to ξ and η does not vanish in the domain $\bar{H}$. In virtue of (3) the Jacobian is everywhere continuous, and of fixed sign.

From (2) and (3) it follows that, if $(\xi + \Delta\xi, \eta + \Delta\eta)$, (ξ, η) be two points of $\bar{H}$, and $(x + \Delta x, y + \Delta y)$, (x, y) the corresponding points of H, then

$$\Delta x = \left(\frac{\partial f_1}{\partial \xi} + \theta_1\right)\Delta\xi + \left(\frac{\partial f_1}{\partial \eta} + \theta_2\right)\Delta\eta,$$

$$\Delta y = \left(\frac{\partial f_2}{\partial \xi} + \theta_3\right)\Delta\xi + \left(\frac{\partial f_2}{\partial \eta} + \theta_4\right)\Delta\eta,$$

where θ_1, θ_2, θ_3, θ_4 converge to zero as $\Delta\xi$, $\Delta\eta$ do so, and (see § 237) uniformly for all points (ξ, η) in any closed domain contained in $\bar{H}$. On solving these equations, we find

$$\Delta\xi = \frac{\left(\frac{\partial f_2}{\partial \eta} + \theta_4\right)\Delta x - \left(\frac{\partial f_1}{\partial \eta} + \theta_2\right)\Delta y}{J + \alpha_1},$$

with a similar expression for $\Delta\eta$, where J denotes the Jacobian

$$\frac{\partial(f_1, f_2)}{\partial(\xi, \eta)},$$

and α_1 is a function of θ_1, θ_2, θ_3, θ_4 which converges with them to zero. Since, by (4), J never vanishes, it follows from these equations that $\Delta\xi$, $\Delta\eta$ converge to zero with Δx, Δy, and thus that the functions $\phi_1(x, y)$, $\phi_2(x, y)$ are both continuous functions.

The partial differential coefficients

$$\frac{\partial \phi_1}{\partial x} = \frac{\partial f_2}{\partial \eta}\Big/ J, \quad \frac{\partial \phi_1}{\partial y}, \quad \frac{\partial \phi_2}{\partial x}, \quad \frac{\partial \phi_2}{\partial y}$$

are also continuous in H; and therefore

$$\Delta\xi = \left(\frac{\partial \phi_1}{\partial x} + \chi_1\right)\Delta x + \left(\frac{\partial \phi_1}{\partial y} + \chi_2\right)\Delta y,$$

$$\Delta\eta = \left(\frac{\partial \phi_2}{\partial x} + \chi_3\right)\Delta x + \left(\frac{\partial \phi_2}{\partial y} + \chi_4\right)\Delta y,$$

where χ_1, χ_2, χ_3, χ_4 converge to zero with Δx, Δy, and uniformly so for all points (x, y) in H.

Corresponding to any closed set h, of zero content, contained in H, there is a closed set $\bar{h}$, of zero content, contained in $\bar{H}$. It is clear, from the continuity of the functions which define the transformation, that a limiting point of a sequence of points in H corresponds to the limiting point of the corresponding sequence in $\overline{H}$; and thus $\overline{h}$ is closed, since h is so.

Writing $\Delta\xi = L\Delta x + M\Delta y, \quad \Delta\eta = L'\Delta x + M'\Delta y,$
it follows, since

$$\frac{(\Delta\xi)^2 + (\Delta\eta)^2}{(\Delta x)^2 + (\Delta y)^2} \leqq L^2 + L'^2 + M^2 + M'^2 + 2\,|\,LM + L'M'\,|,$$

where $L, L', \ldots$ converge uniformly to

$$\frac{\partial\phi_1}{\partial x}, \quad \frac{\partial\phi_2}{\partial x}, \quad \ldots,$$

that, if $|\Delta x|, |\Delta y|$ be both restricted to be less than a fixed positive number ϵ, the ratio

$$\frac{(\Delta\xi)^2 + (\Delta\eta)^2}{(\Delta x)^2 + (\Delta y)^2}$$

has a finite upper limit Λ^2, for the whole domain H. Now let the points of h be enclosed in a finite number of circles, the radii of which are all $< \epsilon$; it then follows that the points of $\overline{h}$ can be enclosed in a finite number of circles of which the radii are all less than $\epsilon\Lambda$. The sum of the areas of these circles on the (ξ, η) plane, which contain in their interiors all the points of $\overline{h}$, has to the sum of the areas of the circles on the (x, y) plane, which enclose all the points of h, a ratio less than Λ^2. Since the sum of the latter circles can be taken to be arbitrarily small, it follows that the points of $\overline{h}$ can all be enclosed in a finite number of circles the sum of whose areas is arbitrarily small. Therefore $\overline{h}$ has the content zero.

326. Let $f(x, y)$ be a limited function, defined for all points of a closed connex domain G contained in H, the frontier of G having content zero; and let $f(x, y)$ be integrable in G. If x, y be expressed in terms of ξ, η by the relations

$$x = f_1(\xi, \eta), \quad y = f_2(\xi, \eta),$$

which satisfy the conditions of § 325, then, corresponding to $f(x, y)$ in G, we have a function $F(\xi, \eta)$ in the domain $\overline{G}$, contained in $\overline{H}$, which corresponds to G. The frontier of $\overline{G}$, corresponding to the frontier of G, has also the content zero. A point of discontinuity of $f(x, y)$ in the (x, y) plane corresponds to a point of discontinuity of $F(\xi, \eta)$ in the (ξ, η) plane, the measures of discontinuity at the corresponding points being the same. Since those points of (x, y) at which the saltus of $f(x, y)$ is $\geqq k$ form a closed set of zero content, it follows that the points of (ξ, η) at which the saltus of $F(\xi, \eta)$ is $\geqq k$ form also a closed set of zero content; and therefore $F(\xi, \eta)$ is integrable in $\overline{G}$.

In order to transform $\int f(x, y)\,(dx dy)$, taken throughout G, into an integral taken throughout $\overline{G}$, it is convenient to make use of an intermediate transformation* $x=\psi(u_1, u_2)$, $y=u_2$, followed by the transformation $u_1=\xi$, $u_2=f_2(\xi, \eta)$; the function $\psi(u_1, u_2)$ being such that

$$\psi(u_1, u_2)=f_1(\xi, \eta).$$

It is easy to see that each of these transformations satisfies the conditions of § 325.

Since $f(x, y)$ is integrable in G, we may, in accordance with the result of § 314, replace the double integral

$$\int f(x, y)\,(dxdy)$$

by the repeated integral $$\int dy \overline{\int} f(x, y)\,dx,$$

or by $$\int dy \underline{\int} f(x, y)\,dx.$$

Applying the transformation $x=\psi(u_1, u_2)$ to the upper and lower integrals

$$\overline{\int} f(x, y)\,dx, \qquad \underline{\int} f(x, y)\,dx,$$

these may, in virtue of the theorem of § 306, be transformed into the single upper and lower integrals,

$$\overline{\int} \phi(u_1, u_2)\frac{\partial x}{\partial u_1}\,du_1, \qquad \underline{\int} \phi(u_1, u_2)\frac{\partial x}{\partial u_1}\,du_1,$$

where $\phi(u_1, u_2)$ represents the function of u_1, u_2 which corresponds to $f(x, y)$. We thus have

$$\begin{aligned}\int f(x, y)\,(dxdy) &= \int du_2 \overline{\int} \phi(u_1, u_2)\frac{\partial x}{\partial u_1}\,du_1\\ &= \int du_2 \underline{\int} \phi(u_1, u_2)\frac{\partial x}{\partial u_1}\,du_1\\ &= \int \phi(u_1, u_2)\frac{\partial x}{\partial u_1}\,(du_1du_2),\end{aligned}$$

the double integral being taken through the domain in the plane (u_1, u_2), which corresponds to G in (x, y).

Since $\dfrac{\partial x}{\partial u_1}$ is the Jacobian of (x, y) with respect to (u_1, u_2), we have by a known theorem

$$J=\frac{\partial x}{\partial u_1}\cdot\frac{\partial(u_1, u_2)}{\partial(\xi, \eta)};$$

* This method is employed in the general case of multiple integrals by Pierpont; see his paper "On multiple integrals," *Trans. Amer. Math. Soc.*, vol. VI, p. 432. It is, however, there assumed that $f(x, y)$ is integrable with respect to x for each value of y: but this is unnecessary.

hence, since J never vanishes,

$$\frac{\partial x}{\partial u_1} \text{ and } \frac{\partial (u_1, u_2)}{\partial (\xi, \eta)}$$

also never vanish.

Applying the same method of transformation to

$$\int \phi (u_1, u_2) \frac{\partial x}{\partial u_1} (du_1 du_2),$$

where $$u_1 = \xi, \quad u_2 = f_2(\xi, \eta),$$

we have $$\int f(x, y)(dxdy) = \int F(\xi, \eta) \frac{\partial x}{\partial u_1} \frac{\partial u_2}{\partial \eta} (d\xi d\eta),$$

where $$\frac{\partial u_2}{\partial \eta} = \frac{\partial (u_1, u_2)}{\partial (\xi, \eta)};$$

hence finally we obtain the formula

$$\int f(x, y)(dxdy) = \int F(\xi, \eta) J (d\xi d\eta);$$

which is the formula of transformation of the integral of $f(x, y)$ throughout G into an integral through $\overline{G}$.

It has been assumed that J has a fixed sign throughout the domain of integration. If now this sign be negative, the product $\Delta\xi\Delta\eta$, in $J\Delta\xi\Delta\eta$, which corresponds to $\Delta x \Delta y$ in the plane of (x, y), must be accounted negative, when $\Delta x \Delta y$ is positive. It is however more convenient to consider $\Delta\xi\Delta\eta$ as essentially positive, otherwise the measure of a set of points in the (ξ, η) plane would have to be reckoned as negative. Adopting this convention, we write $|J|\Delta\xi\Delta\eta$ instead of $Jd\xi d\eta$; and therefore the formula of transformation will be written in the form

$$\int f(x, y)(dxdy) = \int F(\xi, \eta) |J| (d\xi d\eta).$$

327. Let us now assume that at certain points of G, which form a set L of zero content, either (1) $f(x, y)$ has an infinite discontinuity, or (2) one or more of the partial differential coefficients

$$\frac{\partial f_1}{\partial \xi}, \quad \frac{\partial f_1}{\partial \eta}, \quad \frac{\partial f_2}{\partial \xi}, \quad \frac{\partial f_2}{\partial \eta}$$

does not exist, or is discontinuous, or (3) the Jacobian J vanishes. In case J be positive over a part of G, and negative over another part, it is convenient to divide the double integral into two portions, taken over these two parts of G respectively, and to transform these two portions separately. It will accordingly be assumed that J never actually changes its sign in the domain G, although it may vanish at the points of the part L of G. We may denote by $\overline{L}$ the set of points on the (ξ, η) plane which corresponds to L: it will be assumed that $\overline{L}$ has zero content. It will be shewn that, if one of the two integrals

$$\int f(x, y)\,(dxdy),$$

taken over G, and

$$\int F(\xi, \eta)\,|J|\,(d\xi d\eta),$$

taken over $\overline{G}$, exists as an absolutely convergent improper integral, or as a proper integral, then the other one exists, and the two have the same value.

Let us assume that $\int_G f(x, y)\,(dxdy)$ exists: it will then be sufficient, in order to establish the existence of the other integral, and its equality with the first, to shew that, for any domain $\overline{G}_1$ contained in $\overline{G}$ and itself containing no points of $\overline{L}$ either in its interior or on its frontier (which frontier is to be taken to be of zero measure), the condition

$$\left|\int_G f(x, y)\,(dxdy) - \int_{\overline{G}_1} F(\xi, \eta)\,|J|\,(d\xi d\eta)\right| < \eta$$

is satisfied, provided that $m(\overline{G}) - m(\overline{G}_1)$ be less than some fixed finite number dependent on η.

A domain g interior to G, and containing in its interior and on its frontier no points of L, can be found such that

$$\left|\int_G f(x, y)\,(dxdy) - \int_g f(x, y)\,(dxdy)\right| < \epsilon.$$

If h be any domain contained in g, such that $m(g) - m(h)$ is sufficiently small, we have

$$\left|\int_g f(x, y)\,(dxdy) - \int_h f(x, y)\,(dxdy)\right| < \epsilon;$$

and therefore

$$\left|\int_G f(x, y)\,(dxdy) - \int_h f(x, y)\,(dxdy)\right| < 2\epsilon.$$

Now let k be a domain interior to G, containing in its interior and on its frontier no points of L, and containing h, then

$$\left|\int_G f(x, y)\,(dxdy) - \int_k f(x, y)\,(dxdy)\right| < 2\epsilon.$$

For let p denote the domain obtained by taking the two domains g and k together, then

$$\left|\int_G f(x, y)\,(dxdy) - \int_p f(x, y)\,(dxdy)\right| < \epsilon,$$

and

$$\left|\int_p f(x, y)\,(dxdy) - \int_k f(x, y)\,(dxdy)\right| < \epsilon,$$

and by combining these inequalities the result follows.

We have now

$$\int_{\overline{G}_1} F(\xi, \eta)\,|J|\,(d\xi d\eta) = \int_{\overline{U}} F(\xi, \eta)\,|J|\,(d\xi d\eta) - \int_{\overline{V}} F(\xi, \eta)\,|J|\,(d\xi d\eta),$$

where $\overline{U}$ is the domain formed by taking all points which belong to one or both of the domains $\overline{G}_1$ and $\overline{h}$; and $\overline{V}$ consists of those points which belong to $\overline{h}$ but not to $\overline{G}_1$.

Now $\overline{U}$ corresponds to a domain in the (x, y) plane which contains h, and which domain may be taken to be identical with k; therefore

$$\int_{\overline{U}} F(\xi, \eta)\,|J|\,(d\xi d\eta) \text{ differs from } \int_G f(x, y)\,(dxdy)$$

by a number numerically less than 2ϵ. Again

$$\left|\int_{\overline{V}} F(\xi, \eta)\,|J|\,(d\xi d\eta)\right| < \mu\,\{m(\overline{G}) - m(\overline{G}_1)\}$$

where μ is the upper limit of $|F(\xi, \eta)\,J|$ in the domain $\overline{V}$ obtained by removing from $\overline{h}$ those points which belong to $\overline{G}_1$.

We thus have

$$\left|\int_G f(x, y)\,(dxdy) - \int_{\overline{G}_1} F(\xi, \eta)\,|J|\,(d\xi d\eta)\right| < 2\epsilon + \mu\,\{m(\overline{G}) - m(\overline{G}_1)\}.$$

Now let ϵ be fixed so that it is $< \frac{1}{4}\eta$, then $\overline{h}$ is fixed, so that μ cannot exceed a fixed finite number μ_1. If then $\overline{G}_1$ be so chosen that

$$m(\overline{G}) - m(\overline{G}_1) < \frac{\eta}{2\mu_1},$$

the inequality

$$\left|\int_G f(x, y)\,(dxdy) - \int_{\overline{G}_1} F(\xi, \eta)\,|J|\,(d\xi d\eta)\right| < \eta$$

will be satisfied. Therefore it follows that

$$\int_{\overline{G}} F(\xi, \eta)\,|J|\,(d\xi d\eta)$$

exists, and is equal to

$$\int_G f(x, y)\,(dxdy).$$

328. This method of transformation may be extended to the case in which one of the domains G, $\overline{G}$ is infinite, or to the case in which both are infinite. It can be shewn that, if either of the integrals

$$\int_G f(x, y)\,(dxdy), \qquad \int_{\overline{G}} F(\xi, \eta)\,|J|\,(d\xi d\eta)$$

exist, the definition of § 323 being applied when G or $\overline{G}$ is infinite, then the other integral also exists, and the two integrals are equal. The proof can be given by slightly modifying the procedure of § 327.

There may be a set of points of zero content, in the domain of G, such that the corresponding values of ξ, η are infinite, or such that one of them is infinite. This set now takes the place of the set L. Whether G be finite or infinite, the finite, or infinite, domain h contained in G may be so fixed as to exclude all points which correspond to infinite values of ξ or η. The domain k including h, and containing no points which correspond to infinite values of ξ and η, may then be fixed as before, and will satisfy the condition

$$\left|\int_G f(x, y)(dxdy) - \int_k f(x, y)(dxdy)\right| < 2\epsilon,$$

it being assumed that the integral of $f(x, y)$ over G exists.

The finite domain $\bar{h}$ contains all points of $\bar{G}$ of which the distance from the origin is less than some number R depending on the domain $G-h$, which contains in its interior all points (x, y) that correspond to infinite values of ξ or η or of both. The same statement holds for $\bar{k}$, which contains $\bar{h}$. When the finite domain $\bar{G}_1$ is such that the condition

$$\int_{\bar{V}} F(\xi, \eta)\,|J|\,(d\xi d\eta) < \tfrac{1}{2}\eta$$

is satisfied (and, in order that this may be the case, $\bar{G}_1$ must certainly contain all points of $\bar{G}$ whose distance from the origin is less than some fixed number $R_1 \leq R$), we have as before

$$\left|\int_G f(x, y)(dxdy) - \int_{\bar{G}_1} F(\xi, \eta)\,|J|\,(d\xi d\eta)\right| < \eta;$$

and as η is an arbitrarily fixed number, we thus see that

$$\int_{\bar{G}} F(\xi, \eta)\,|J|\,(d\xi d\eta)$$

exists and is equal to

$$\int_G f(x, y)(dxdy).$$

CHAPTER VI.

FUNCTIONS DEFINED BY SEQUENCES.

329. LET us suppose that $u_1, u_2, u_3, \ldots u_n, \ldots$ is an unending sequence of numbers, so that u_n has for each value of n a definite numerical value, assigned by means of a prescribed rule or set of rules. Let the sums u_1, $u_1 + u_2$, $u_1 + u_2 + u_3, \ldots u_1 + u_2 + \ldots + u_n, \ldots$ be denoted by $s_1, s_2, s_3, \ldots s_n, \ldots$, and let us consider the aggregate $(s_1, s_2, \ldots, s_n \ldots)$. If this aggregate be a convergent one, in the sense described in § 28, it has a limit s_ω or s, which is said to be *the limiting sum* of the infinite series $u_1 + u_2 + \ldots + u_n + \ldots$, in which case the series is said to be convergent.

The condition that the sequence $(s_1, s_2, s_3, \ldots s_n, \ldots)$ be convergent is that, corresponding to each arbitrarily chosen positive number ϵ, a value of n can be found such that $|s_{n+m} - s_n| < \epsilon$, for $m = 1, 2, 3, \ldots$. This is then the condition that the infinite series $u_1 + u_2 + \ldots + u_n + \ldots$ may be convergent.

The difference $s_{n+m} - s_n \equiv u_{n+1} + u_{n+2} + \ldots + u_{n+m}$ is called a *partial remainder* of the infinite series, and may be denoted by $R_{n,m}$. Thus the condition of convergence of the infinite series may be stated in the form, that, *corresponding to each arbitrarily chosen positive number ϵ, a value of n can be found such that all the partial remainders $R_{n,1}, R_{n,2}, \ldots R_{n,m}, \ldots$ are numerically less than ϵ.*

Since $R_{n-1,1} = u_n$, it is seen to be a necessary, but not a sufficient, condition for the convergence of the series, that $|u_n|$ be arbitrarily small when n is sufficiently great; this condition may be written in the form $\lim_{n=\infty} u_n = 0$.

If the series $u_1 + u_2 + \ldots + u_n + \ldots$ be convergent, then, for any value of n, the series $u_{n+1} + u_{n+2} + \ldots$ is also convergent, and has, in the sense defined above, a limiting sum which may be denoted by R_n. This limit is called the *remainder after n terms* of the original convergent series; thus $s = s_n + R_n$.

It is clear that, the given series being convergent, the sequence $R_1, R_2, \ldots R_n, \ldots$ is also convergent, and that its limit is zero. That this may be the case has sometimes been given as the necessary and sufficient condition for the convergence of the given series; such a statement of the condition is, however, circular, because the existence of the numbers R_n cannot be assumed unless the given series is already known to be convergent.

It is important to observe that the number s has not been defined as the sum of the infinite series $u_1 + u_2 + \ldots + u_n + \ldots$; for that would have implied the completion of an indefinite series of operations of addition: but, conversely, the limiting sum, or simply the sum, of the infinite series has been defined to be that number s which was itself defined, as in § 28, by means of a convergent sequence.

330. If all the terms of the series $u_1 + u_2 + \ldots + u_n + \ldots$ be positive, the numbers of the sequence $s_1, s_2, \ldots s_n, \ldots$ continually increase; and hence it is a sufficient condition for the convergence of the series that a number K exist such that $s_n < K$ for every value of n; for the numbers $s_1, s_2, \ldots s_n, \ldots$ then have an upper limit s.

If the terms of the series $u_1 + u_2 + u_3 + \ldots + u_n + \ldots$ be of alternate signs, or if this be the case from and after some fixed term, then the necessary and sufficient conditions for the convergence of the series are that $|u_n|$ should continually diminish as n is increased, and that $\lim_{n=\infty} u_n = 0$. For, in this case $|s_{n+m} - s_n| < |u_{n+1}|$, for every value of m, from and after some fixed value of n, and $|u_{n+1}|$ is arbitrarily small, if a sufficiently great value of n be chosen; hence the sequence $s_1, s_2, \ldots s_n, \ldots$ is convergent.

If the series $u_1 + u_2 + \ldots + u_n + \ldots$ be convergent, and $a_1, a_2, \ldots a_n, \ldots$ be a sequence of numbers which, from and after some particular value of n, are all positive, and do not increase as n is increased, then the series

$$a_1 u_1 + a_2 u_2 + \ldots + a_n u_n + \ldots$$

is also convergent.

In proving this theorem, it is clear that, without loss of generality, we may suppose all the numbers $a_1, a_2, \ldots$ to be positive and not increasing, since we need only remove a definite number of terms from each series to reduce the general case to this one. We have

$$a_{n+1}u_{n+1} + a_{n+2}u_{n+2} + \ldots + a_{n+m}u_{n+m}$$
$$= a_{n+1}R_n + (a_{n+2} - a_{n+1})R_{n+1} + \ldots + (a_{n+m} - a_{n+m-1})R_{n+m-1} - a_{n+m}R_{n+m}.$$

Since the series Σu is convergent, we can, corresponding to an arbitrarily chosen ϵ, find n such that $R_n, R_{n+1}, R_{n+2}, \ldots$ are all numerically less than ϵ; also $a_{n+2} - a_{n+1}, a_{n+3} - a_{n+2}, \ldots$ are all of the same sign, therefore

$$|a_{n+1}u_{n+1} + a_{n+2}u_{n+2} + \ldots + a_{n+m}u_{n+m}| < a_{n+1}\epsilon + |a_{n+m} - a_{n+1}|\epsilon + a_{n+m}\epsilon$$
$$< 2a_{n+1}\epsilon < 2a_1\epsilon;$$

and thus, from and after a large enough value of n, all the partial remainders of the series Σau are arbitrarily small, and therefore the series is convergent.

It is clear from the preceding proof that the theorem also holds if, from and after some fixed index n, the numbers $a_1, a_2, a_3, \ldots$ do not diminish, but be such that all of them are less than some fixed number.

NON-CONVERGENT ARITHMETIC SERIES.

331. The partial sums $s_1, s_2, \dots s_n, \dots$ of a series $u_1 + u_2 + \dots + u_n + \dots$ may be represented in the usual manner by an enumerable set of points G, on a straight line. The following cases may arise:—

(1) The set G may all lie between two fixed points A, B, and the derivative G' may consist of a single point s; in this case the series is *convergent*, and s is the limiting sum.

(2) There may not exist any two points A, B between which all the points of G lie, and G' also may not exist; in this case $|s_n|$ has no upper limit, and the series is said to be *divergent*.

A divergent series may be such that, from and after some fixed term, all the partial sums are of the same sign and increase without limit. An example of a divergent series of this kind is the series

$$\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} + \dots.$$

A divergent series may be such that, although $|s_n|$ increases without limit, there are an unlimited number of positive elements s_n, and also an unlimited number of negative elements s_n. An example of this class of divergent series is the series $1 - 2 + 3 - 4 + \dots + (2n-1) - 2n + \dots$, for which $s_{2n-1} = n$, $s_{2n} = -n$.

(3) The set G may consist of points all lying between two fixed points A and B, and the derivative G' may consist of more than one point; in this case the series is said to be an *oscillating series*. The set G' may contain a finite, or an infinite, number of points, but it must be a closed set; it consequently has an upper boundary U and a lower boundary L; and these boundaries U, L are called* the *limits of indeterminacy* of the series. It is always possible to find a sequence $(s_{n_1}, s_{n_2}, s_{n_3}, \dots)$ of partial sums, where $n_1 < n_2 < n_3, \dots$, which converges to the point U, and another such sequence which converges to L, or to any point of G' which may be chosen. It thus appears that, by introducing a suitable system of bracketing, according to some norm, the terms of an oscillating series, the series may be converted into a convergent one of which the limiting sum is any chosen point of G', including either limit of indeterminacy. The set G' may be non-dense in the interval (L, U), or it may consist of all the points of that interval, or it may consist of a closed set of the most general type, as described in § 86.

The series $1 - 1 + 1 - 1 + 1 \dots$ has 1 and 0, for the upper and lower limits of indeterminacy; and G' may be regarded as consisting of these two points.

* This term is due to Du Bois Reymond: see his *Antrittsprogramm*, p. 3.

Let $s_1 = \frac{1}{2},\ s_2 = \frac{1}{3},\ s_3 = \frac{1}{4},\ s_4 = \frac{2}{3},\ s_5 = \frac{1}{5},\ s_6 = \frac{2}{4},\ s_7 = \frac{1}{6}, \ldots$

and generally

$$s_{m(m+1)+1} = \frac{1}{2m+2},\quad s_{m(m+1)+2} = \frac{2}{2m+1}, \ldots s_{(m+1)^2} = \frac{m+1}{m+2},$$

$$s_{(m+1)^2+1} = \frac{1}{2m+3},\quad s_{(m+1)^2+2} = \frac{2}{2m+2}, \ldots s_{(m+1)(m+2)} = \frac{m+1}{m+3}.$$

It follows that the series

$$\frac{1}{2} - \frac{1}{2.3} - \frac{1}{3.4} + \frac{5}{3.4} - \frac{7}{3.5} + \ldots$$

has 1 and 0 for the upper and lower limits of indeterminacy. The set G consists of all the rational numbers between 0 and 1; so that G' consists of the whole interval (0, 1). By introducing a properly chosen system of brackets, the series may be converted into one converging to a limiting sum which is any prescribed number of the interval (0, 1).

(4) The derivative G' may exist, but one or both of the points A, B may be absent; in this case also the series is said to be an oscillating series. If the points s_n have no upper boundary, then the upper limit of indeterminacy is said to be $+\infty$; and if they have no lower boundary, the lower limit of indeterminacy is said to be $-\infty$. In this case the series may be made to diverge, by introducing a properly chosen system of brackets, or, on the other hand, it may be made to converge to any point of G'.

It should be observed that oscillating series are frequently included in the term divergent series.

For example, a series* may be constructed which oscillates between infinite limits of indeterminacy, but which, by introducing a suitable system of brackets in accordance with a norm, may be made to converge to any prescribed number whatever.

If $x' = \frac{2x-1}{\sqrt{x(1-x)}}$, where the positive sign is ascribed to the radical, the points x of the interval (0, 1) have a (1, 1) correspondence with the points x' of the unlimited straight line $(-\infty, \infty)$. It is clear that a set of points $\{x\}$ in the interval (0, 1) corresponds to a set $\{x'\}$ in $(-\infty, \infty)$, the relation of order being conserved in the correspondence. Further, a limiting point of the one set corresponds to a limiting point of the other set. The rational points of the interval (0, 1) of x correspond to a set of points x' everywhere-dense in $(-\infty, \infty)$. This method of transformation may be applied to the series obtained in (3), which oscillates between the limits of indeterminacy 0, 1, and which can be made, by introducing suitable brackets, to converge to any prescribed number in the interval (0, 1).

* See Hobson, *Proc. Lond. Math. Soc.*, ser. 2, vol. III, p. 50.

We find that

$$s_1' = 0, \quad s_2' = -\frac{1}{\sqrt{2}}, \quad s_3' = -\frac{2}{\sqrt{3}}, \quad s_4' = \frac{1}{\sqrt{2}}, \quad s_5' = -\frac{3}{2},$$

$$s_6' = 0, \quad s_7' = -\frac{4}{\sqrt{5}}, \quad s_8' = -\frac{1}{\sqrt{6}}, \quad s_9' = \frac{2}{\sqrt{3}}, \quad s_{10}' = -\frac{5}{\sqrt{6}};$$

and generally

$$s'_{m(m+1)+1} = -\frac{2m}{\sqrt{(2m+1)}}, \quad s'_{m(m+1)+2} = -\frac{2m-3}{\sqrt{\{2(2m-1)\}}}, \ldots$$

$$s'_{(m+1)^2} = \frac{m}{\sqrt{(m+1)}}, \quad s'_{(m+1)^2+1} = -\frac{2m+1}{\sqrt{(2m+2)}},$$

$$s'_{(m+1)(m+2)} = \frac{m-1}{\sqrt{\{2(m+1)\}}}.$$

Therefore the series

$$-\frac{1}{\sqrt{2}} + \left(\frac{1}{\sqrt{2}} - \frac{2}{\sqrt{3}}\right) + \left(\frac{2}{\sqrt{3}} + \frac{1}{\sqrt{2}}\right) - \left(\frac{3}{2} + \frac{1}{\sqrt{2}}\right) + \frac{3}{2} - \frac{4}{\sqrt{5}} + \ldots$$

has the required character: it may be made to converge to any assigned number whatever, by suitably bracketing the terms together in accordance with a norm, and amalgamating the terms in each bracket.

ABSOLUTELY CONVERGENT, AND CONDITIONALLY CONVERGENT, SERIES.

332. Let us suppose the terms of a convergent arithmetical series to be all positive. The order of the terms in the series is defined by the norm which defines the series.

If now a new series be defined by another norm, and be such that any assigned term in either series is identical with a definite term in the other series, then the new series is said to be obtained by rearranging the terms of the original series; and the two series are conventionally regarded as identical with one another. It can be shewn that the new series converges to the sum of the original series.

Let $(s_1, s_2, \ldots s_n, \ldots)$, $(s_1', s_2', \ldots s'_{n'}, \ldots)$ be the aggregates of partial sums of the two series, and let s be the sum of the given series defined by the first of these aggregates. If ϵ be an arbitrarily chosen positive number, n can be determined so that $s - s_n < \epsilon$; and then, since all the terms of the series are positive, it follows that $s - s_{n+m} < \epsilon$, for $m = 1, 2, 3, \ldots$.

A number n' can be determined such that the first n terms of the first series all occur in the first n' terms of the second series; and therefore $s_n < s'_{n'}$. Again a number n'' can be determined so that the first n' terms of the second series all occur in the first n'' terms of the first series; then we have $s_n < s'_{n'} < s_{n''}$. Since $s - s_n$, $s - s_{n''}$ are both $< \epsilon$, it follows that $s - s'_{n'} < \epsilon$;

and this clearly holds for all values of n' greater than the one employed. Since ϵ is arbitrary, it has thus been shewn that the second aggregate of partial sums converges to s; and therefore *the sum of a convergent series whose terms have all the same sign is unaltered, if the order of the terms be altered in accordance with some norm, such that each term of the original series has a definite place in the new series.*

333. Next let us suppose the arithmetical series $u_1 + u_2 + u_3 + \dots$ to have both positive and negative terms, each indefinitely great in number. Let the positive terms, in the order in which they occur, be $a_1, a_2, a_3, \dots$ and the negative terms, in the order in which they occur, be $-b_1, -b_2, -b_3, \dots$; and consider the two series

$$a_1 + a_2 + a_3 + \dots, \quad \dots\dots\dots\dots\dots\dots(1)$$

$$b_1 + b_2 + b_3 + \dots. \quad \dots\dots\dots\dots\dots\dots(2)$$

Denoting by σ_m, $\sigma'_{m'}$ the sums of m and m' terms respectively of these series, we see that, if in the first n terms of the given series there are m positive and m' negative terms, then

$$s_n = \sigma_m - \sigma'_{m'}.$$

If both the series (1), (2) be convergent, then σ_m, $\sigma'_{m'}$ have finite limits, and the given series is itself convergent, its sum being independent of the arrangement of the positive and negative terms, provided only that each term of the original series occurs in the series obtained by the rearrangement of the order of the terms.

The series $u_1 + u_2 + u_3 + \dots$ is said to be absolutely convergent, provided $|u_1| + |u_2| + |u_3| + \dots$ be convergent, which is the case when both the series (1), (2) formed respectively by the positive and by the negative terms, are convergent. *The sum of an absolutely convergent series is thus unaltered by a rearrangement of the terms, in accordance with some norm.*

If one of the two series (1), (2) be convergent, and the other be not convergent, the given series is not convergent.

If both the series (1) and (2) be divergent, it may happen that the given series itself is convergent; in this case $\sigma_m - \sigma'_{m'}$ has a definite limit, although σ_m, $\sigma'_{m'}$ have no limits.

If the series $u_1 + u_2 + \dots$ be convergent, whilst the series $|u_1| + |u_2| + \dots$ is not convergent, then the given series is said to be conditionally convergent.*

It will be seen that the order of the terms in a conditionally convergent series cannot in general be altered without affecting the sum of the series, or of possibly rendering it no longer convergent.

* By Stokes the term accidentally convergent is used; by many writers such series are spoken of as semi-convergent, but this term is also used in quite another connection.

334. It will now be shewn that *the terms of a conditionally convergent series can always be so rearranged in accordance with a norm, that* (1) *the new series is convergent, and has as sum an arbitrarily given number; or* (2) *so that the new series is divergent; or* (3) *that the new series oscillates between arbitrarily given limits; moreover each of these rearrangements may be made in an indefinitely great number of ways.*

Let $k_1, k_2, k_3, \dots k_n, \dots$ be a sequence of positive numbers, assigned in accordance with any prescribed law, and which are such that no one is less than the preceding one. Take p_1 terms of the series (1), so that $a_1 + a_2 + \dots + a_{p_1}$ or σ_{p_1}, is greater than k_1, whilst $\sigma_{p_1-1} \leqq k_1$; next take q_1 terms of the series (2) such that $a_1 + a_2 + \dots + a_{p_1} - b_1 - b_2 - \dots - b_{q_1}$, or $\sigma_{p_1} - \sigma_{q_1}'$, is less than k_1, whilst $\sigma_{p_1} - \sigma'_{q_1-1} \geqq k_1$. Next take p_2 more terms of the series (1) so that $\sigma_{p_1} - \sigma_{q_1}' + (a_{p_1+1} + a_{p_1+2} + \dots + a_{p_1+p_2})$ is greater than k_2, whilst

$$\sigma_{p_1} - \sigma_{q_1}' + (a_{p_1+1} + a_{p_1+2} + \dots + a_{p_1+p_2-1}) \leqq k_2:$$

then take q_2 more terms of the series (2), such that

$$\sigma_{p_1} - \sigma_{q_1}' + (\sigma_{p_1+p_2} - \sigma_{p_1}) - (\sigma'_{q_1+q_2} - \sigma_{q_1}') < k_2,$$

whilst
$$\sigma_{p_1} - \sigma_{q_1}' + (\sigma_{p_1+p_2} - \sigma_{p_1}) - (\sigma'_{q_1+q_2-1} - \sigma_{q_1}') \geqq k_2.$$

By proceeding in this manner, we define an arrangement of a number of terms of the given series, which is such that

$$\begin{aligned}(a_1 + a_2 + \dots + a_{p_1}) &- (b_1 + b_2 + \dots + b_{q_1}) \\ &+ (a_{p_1+1} + \dots + a_{p_1+p_2}) - (b_{q_1+1} + \dots + b_{q_1+q_2}) \\ &+ \dots \\ &+ (a_{p_1+p_2+\dots+p_{n-1}+1} + \dots + a_{p_1+p_2+\dots+p_n}) - (b_{q_1+\dots+q_{n-1}+1} + \dots + b_{q_1+\dots+q_n}) \quad (3)\end{aligned}$$

is less than k_n, whilst, if we leave out the last bracket, the expression is greater than k_n. It will be observed that none of the numbers $p_1, p_2, \dots p_n$, $q_1, q_2, \dots q_n$ can be zero. The expression we have obtained differs from k_n by less than $b_{q_1+q_2+\dots+q_n}$; and if in the expression we leave out the last bracket, the resulting expression differs from k_n by less than $a_{p_1+p_2+\dots+p_n}$; also, since the given series is convergent, both the numbers $a_{p_1+p_2+\dots+p_n}$, $b_{q_1+q_2+\dots+q_n}$ are as small as we please, if n be taken sufficiently large. Now suppose $(k_1, k_2, \dots k_n, \dots)$ to be a convergent sequence which defines the number k, arbitrarily given, then, taking n large enough, $k_n, k_{n+1}, \dots$ all differ from k by less than an arbitrarily small number; hence a number n can be found such that, for it and for all succeeding integers, the series we have found differs by an arbitrarily small number from k, even when we suppress the last bracket. We have thus assigned, by means of a norm, an arrangement of the terms of the given series so that the new series, so defined, converges to the arbitrarily given number k; and it is clear that this may be done in an indefinitely great number of ways, since there are an indefinitely great number of convergent sequences which define the same number k.

Next suppose the sequence $k_1, k_2, \dots k_n, \dots$ increases without limit; then we have defined an arrangement of the terms of the given series which makes the new series divergent. It is clear that the same method would have been applicable if we had taken the numbers $k_1, k_2, \dots$ all negative and numerically increasing.

It remains to be shewn that the terms of the series can be arranged so that the sum of the new series oscillates between two arbitrarily given numbers k, k' $(k > k')$. We may suppose without loss of generality that k, k' are both positive.

Let $(k_1, k_2, \dots k_n \dots)$ be an aggregate of increasing numbers which defines the number k, and $(k_1', k_2', \dots k_n' \dots)$ another such aggregate which defines k'; where we may suppose that, for all values of n, $k_n > k_n'$.

Choose p_1 so that $\sigma_{p_1} > k_1$ whilst $\sigma_{p_1-1} \leqq k_1$; next choose q_1 so that $\sigma_{p_1} - \sigma_{q_1}' < k_1'$, whilst $\sigma_{p_1} - \sigma'_{q_1-1} \geqq k_1'$; then choose p_2 so that

$$\sigma_{p_1} - \sigma_{q_1}' + (\sigma_{p_1+p_2} - \sigma_{p_1}) > k_2,$$

whilst
$$\sigma_{p_1} - \sigma_{q_1}' + (\sigma_{p_1+p_2-1} - \sigma_{p_1}) \leqq k_2;$$

then choose q_2 so that

$$\sigma_{p_1} - \sigma_{q_1}' + (\sigma_{p_1+p_2} - \sigma_{p_1}) - (\sigma'_{q_1+q_2} - \sigma_{q_1}') < k_2',$$

whilst
$$\sigma_{p_1} - \sigma_{q_1}' + (\sigma_{p_1+p_2} - \sigma_{p_1}) - (\sigma'_{q_1+q_2-1} - \sigma_{q_1}') \geqq k_2'.$$

Proceeding in this manner, we define a series of the form (3) whose sum is $< k_n'$, but differs from k_n' by less than $b_{q_1+q_2+\dots+q_n}$; and moreover, if the last bracket be suppressed, the sum is then $> k_n$, but differs from k_n by less than $a_{p_1+p_2+\dots+p_n}$. It is now clear that an arrangement of the terms of the given series has been assigned, such that the resulting series oscillates between k and k'; and this can, as before, be done in an indefinitely great number of ways.

A special case of this general theorem as to the nature of the new series obtained by rearranging the terms of a conditionally convergent series in accordance with some norm, is that in which $a_1 = b_1$, $a_2 = b_2$, $\dots a_n = b_n \dots$; so that the original series is $a_1 - a_1 + a_2 - a_2 + a_3 - a_3 + \dots$, which, provided $\lim_{n=\infty} a_n = 0$, has zero for its sum. It thus appears that, from the terms of a divergent series

$$a_1 + a_2 + \dots + a_n + \dots,$$

which is such that a_n is arbitrarily small when n is taken large enough, we can, in an indefinitely great number of ways, construct series which are convergent and have a given sum, are divergent, or oscillate between given limits, by taking the various series of the form

$$a_1 + a_2 + \dots + a_{p_1} - a_1 - a_2 - \dots - a_{q_1} + a_{p_1+1} + a_{p_1+2} + \dots + a_{p_1+p_2} - a_{q_1+1} - a_{q_1+2} - \dots - a_{q_1+q_2} + \dots,$$

the numbers $p_1, p_2, \dots q_1, q_2, \dots$ being assigned in the manner explained above.

EXAMPLES.

1. The series

$$1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\dots+\frac{1}{2n-1}-\frac{1}{2n}+\dots$$

is conditionally convergent, its limiting sum being $\log_e 2$. The series

$$1+\frac{1}{3}-\frac{1}{2}+\frac{1}{5}+\frac{1}{7}-\frac{1}{4}+\dots,$$

which is obtained by systematically rearranging the terms of the first series, converges to $\frac{3}{2}\log_e 2$.

For we find that

$$s_{4n}=\sum_{m=1}^{m=n}\left(\frac{1}{4m-3}-\frac{1}{4m-2}+\frac{1}{4m-1}-\frac{1}{4m}\right),$$

and that, for the second series

$$s'_{3n}=\sum_{m=1}^{m=n}\left(\frac{1}{4m-3}+\frac{1}{4m-1}-\frac{1}{2m}\right);$$

therefore

$$s'_{3n}-s_{4n}=\sum_{m=1}^{m=n}\left(\frac{1}{4m-2}-\frac{1}{4m}\right)=\frac{1}{2}s_{2n}.$$

Since s_{4n}, s_{2n} both converge to $\log_e 2$, as n is indefinitely increased, it follows that s'_{3n} converges to $\frac{3}{2}\log_e 2$.

2. By* rearranging the terms of the convergent series

$$1-1+\frac{1}{2}-\frac{1}{2}+\frac{1}{3}-\frac{1}{3}+\dots+\frac{1}{n}-\frac{1}{n}+\dots,$$

we obtain the series

$$1+\frac{1}{2}-1+\frac{1}{3}+\frac{1}{4}-\frac{1}{2}+\frac{1}{5}+\frac{1}{6}-\frac{1}{3}+\dots-\frac{1}{n-1}+\frac{1}{2n-1}+\frac{1}{2n}-\frac{1}{n}+\dots,$$

$$1+\frac{1}{2}+\frac{1}{3}-1+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}-\frac{1}{2}+\dots-\frac{1}{n-1}+\frac{1}{3n-2}+\frac{1}{3n-1}+\frac{1}{3n}-\frac{1}{n}+\dots,$$

$$1-1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}-\frac{1}{2}+\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}+\frac{1}{9}-\frac{1}{3}+\dots$$

$$-\frac{1}{n-1}+\frac{1}{(n-1)^2+1}+\frac{1}{(n-1)^2+2}+\dots+\frac{1}{n^2}-\frac{1}{n}+\dots.$$

The first of these new series converges to $\log_e 2$, the second to $\log_e 3$, and the third diverges.

SERIES OF TRANSFINITE TYPE.

335. If $s_1, s_2, s_3, \dots s_n, \dots s_\omega, s_{\omega+1}, \dots s_\gamma, \dots$

be a set of numbers each one of which is definite, and in which every index that is less than, or equal to, some number β of the second class, occurs as a suffix, and if the series

$$u_1+u_2+\dots+u_n+\dots+u_\omega+u_{\omega+1}+\dots+u_\gamma+\dots$$

* Dini's *Grundlagen*, p. 133.

be formed, where

$$u_1 = s_1,\ u_2 = s_2 - s_1,\ \dots\ u_n = s_n - s_{n-1}\ \dots\ u_\omega = s_{\omega+1} - s_\omega,$$
$$u_{\omega+1} = s_{\omega+2} - s_{\omega+1},\ \dots\ u_\gamma = s_{\gamma+1} - s_\gamma,\ \dots$$

in which the indices of u include every number less than β, then the series is said to be a convergent series* of type β. If β be a limiting number, the series has no last term, but if β be a non-limiting number, the last term of the series is

$$u_{\beta-1} = s_\beta - s_{\beta-1}.$$

An ordinary infinite series

$$u_1 + u_2 + \dots + u_n + \dots$$

is of type ω. A series

$$u_1 + u_2 + \dots + u_n + \dots + v_1 + v_2 + v_3 + \dots + v_n + \dots$$

is of type $\omega 2$; and a double series

$$\sum_{r=1}^{\infty} \sum_{s=1}^{\infty} a_{rs},$$

$$\begin{array}{l} a_{11} + a_{12} + a_{13} + \dots + a_{1n} + \dots \\ + a_{21} + a_{22} + a_{23} + \dots + a_{2n} + \dots \\ + a_{31} + a_{32} + \dots \\ \dots\dots\dots\dots \\ \dots\dots\dots\dots \\ + a_{n1} + a_{n2} + a_{n3} + \dots \\ \dots\dots\dots\dots\dots \\ \dots\dots\dots\dots\dots \end{array}$$

is of type ω^2, if it be taken in columns successively, or in rows successively; but it is of type ω if the terms are taken diagonally in the form

$$a_{11} + (a_{12} + a_{21}) + (a_{13} + a_{22} + a_{31}) + \dots.$$

Conversely, a series of any type β is convergent if all the sums

$$s_1,\ s_2,\ \dots\ s_n,\ \dots\ s_\omega,\ \dots\ s_\beta$$

be definite numbers.

It is clear that, β being any given number of the second class, any series of the ordinary type ω can have its terms so arranged that the series becomes of type β. For a correspondence can be defined between all the ordinal numbers up to, and including β, and the ordinal numbers of the first class.

Let us now suppose that all the terms of a convergent series

$$u_1 + u_2 + \dots + u_\omega + u_{\omega+1} + \dots + u_\gamma + \dots$$

of type β, are positive, and thus that

$$s_1 < s_2 < s_3,\ \dots\ < s_\omega,\ \dots\ < s_\beta.$$

If we represent the numbers

$$s_1,\ s_2,\ \dots\ s_\omega,\ \dots\ s_\beta$$

* Such series have been investigated in a different manner by Hardy, *Proc. Lond. Math. Soc.*, ser. 2, vol. I.

in the usual manner, by points on a straight line, the terms of the series are represented by a set of intervals

$$(0, s_1), (s_1, s_2), \dots (s_\omega, s_{\omega+1}) \dots$$

on the straight line; each interval abuts on the next; and all the points s_α where α is a limiting number of the second class, are semi-external points of the set of intervals. The end-points and the semi-external points of the set of intervals form an enumerable closed set which has consequently zero content; and it follows, from the theory of the measures of sets of points, that the set of intervals has a measure equal to that of the whole interval $(0, s_\beta)$, which is therefore s_β. Since the measure of an infinite sequence of intervals is equal to the sum of the measures of the intervals, it follows that, if the intervals be arranged in a sequence of type ω, their sum is s_β. The following theorem has thus been established:—

If a series of positive numbers be convergent, and of type β, it will also be convergent when arranged in type ω; also the sums will be the same; and conversely.

We may pass to the consideration of series of type β, of which the terms are not necessarily all positive, but of which the convergence is absolute.

An absolutely convergent series of type β is a series which is convergent when each term is replaced by its modulus.

Let us suppose the intervals constructed as before, which represent the terms of the series

$$|u_1| + |u_2| + \dots + |u_\omega| + |u_{\omega+1}| + \dots + |u_\gamma| + \dots.$$

If we choose out from this set of intervals those which correspond to positive terms of the series

$$u_1 + u_2 + \dots + u_\omega + \dots + u_\gamma + \dots,$$

we have a set of intervals which has a definite finite measure; and the same is true of the set of those intervals which correspond to negative terms of the given series. The given series converges to a sum which is the difference of the measures of these two sets of intervals, and this sum is unaffected by the order in which the intervals are taken in either the positive or the negative component. It has thus been shewn that:—

If a series be absolutely convergent, and of type β, then the series is convergent, and its sum is independent of the type.

An important particular case of this theorem is Cauchy's theorem that *an absolutely convergent double series has the same sum whether the sum be taken by rows or by columns.*

DOUBLE SEQUENCES AND DOUBLE SERIES.

336. A set of numbers $\{s_{mn}\}$, where each of the indices m, n ma positive integral number, and the single number s_{mn} is defined, in ac with some norm, for each pair of values of m and n, is said to form *sequence**.

The numbers of the sequence may be regarded as arranged in r columns, in accordance with the scheme

$$\begin{array}{l} s_{11},\ s_{12},\ s_{13},\ \dots\ s_{1n},\dots \\ s_{21},\ s_{22},\ s_{23},\ \dots\ s_{2n},\dots \\ \dots\dots\dots\dots\dots\dots \\ \dots\dots\dots\dots\dots\dots \\ s_{m1},\ s_{m2},\ s_{m3},\dots\ s_{mn},\dots \\ \dots\dots\dots\dots\dots\dots \\ \dots\dots\dots\dots\dots\dots \end{array}$$

in which no column and no row has a last constituent.

If, for a given double sequence, a number s exist, such that, corres to each arbitrarily fixed positive number ϵ, the condition $|s - s_{m}$ satisfied, for every value of m and n such that $m \geqq p$, $n \geqq p$, where p fixed integer dependent on ϵ, then the double sequence is said to be *co* and the number s is said to be the *limit of the double sequence.*

This is denoted by

$$s = \lim_{m=\infty,\, n=\infty} s_{mn}.$$

The theory of double sequences may be correlated with that g §§ 231, 232, of the double and the repeated limits of a function variables. For, if we assume $x = 1/m$, $y = 1/n$, then the number s_{mn} taken to define the value of a function $f(x, y)$ at the point $x = 1/m$, That the function $f(x, y)$ is not defined for all positive values of x a the neighbourhood of the point $x = 0$, $y = 0$ makes no difference as the validity of the results obtained for a function of two variables. results may now be interpreted as properties of the sequence $\{s_{mn}\}$.

The double limit $\lim\limits_{x=0,\, y=0} f(x, y)$, when it exists, is identic $\lim\limits_{m=\infty,\, n=\infty} s_{mn}$, and the existence of the one double limit implies tha other one.

Corresponding to

$$\overline{\lim_{y=0}}\, f(x, y),\quad \underline{\lim_{y=0}}\, f(x, y),\quad \lim_{y=0} f(x, y),$$

* The theory of double sequences has been treated by Pringsheim, *Münch. Sitzun* vol. XXVIII, 1898; also in *Math. Annalen*, vol. LIII. See also a paper by London in *Math.* vol. LIII, 1900.

the notation $\overline{\lim_{n=\infty}}\, s_{mn}, \quad \underline{\lim_{n=\infty}}\, s_{mn}, \quad \lim_{n=\infty} s_{mn}$

may be employed to denote the upper limit, the lower limit, or the limit of the mth row of the sequence; the limit existing when the upper and lower limits are identical. The notation $\overline{\underline{\lim}}_{n=\infty}\, s_{mn}$ may be used to denote the upper, and the lower, limits, when either is to be taken indifferently. The corresponding notation

$$\overline{\lim_{m=\infty}}\, s_{mn}, \quad \underline{\lim_{m=\infty}}\, s_{mn}, \quad \overline{\underline{\lim}}_{m=\infty}\, s_{mn}, \quad \lim_{m=\infty} s_{mn}$$

is applicable to the nth column of the sequence.

The repeated limits $\lim_{m=\infty} \lim_{n=\infty} s_{mn}$, $\lim_{n=\infty} \lim_{m=\infty} s_{mn}$ correspond precisely to the repeated limits $\lim_{x=0} \lim_{y=0} f(x, y)$, $\lim_{y=0} \lim_{x=0} f(x, y)$ respectively.

The following results are obtained from those in § 232.

The existence of $s \equiv \lim_{m=\infty,\, n=\infty} s_{mn}$ *implies the existence of the repeated limits* $\lim_{m=\infty} \lim_{n=\infty} s_{mn}$, $\lim_{n=\infty} \lim_{m=\infty} s_{mn}$, *which have both the value* s.

The existence of s *is however not a necessary consequence of the existence, and the equality, of the two repeated limits.*

The existence of the repeated limit $\lim_{m=\infty} \lim_{n=\infty} s_{mn}$ *does not necessarily involve that of* $\lim_{n=\infty} s_{mn}$, *as a definite number; but it implies that*

$$\lim_{m=\infty} \overline{\lim_{n=\infty}}\, s_{mn}, \quad \lim_{m=\infty} \underline{\lim_{n=\infty}}\, s_{mn}$$

have one and the same value.

In case the sequence be such that $s_{m'n'} \geqq s_{mn}$, for every value of m', n', such that $m' \geqq m$, $n' \geqq n$, and for every value of m and n, then the sequence is said to be *monotone*. It is also said to be monotone in case the relation $s_{m'n'} \leqq s_{mn}$ is always satisfied.

The following theorem may be easily established:—

If the sequence $\{s_{mn}\}$ *be monotone, then the existence of any one of the three limits* $\lim_{m=\infty,\, n=\infty} s_{mn}$, $\lim_{m=\infty} \lim_{n=\infty} s_{mn}$, $\lim_{n=\infty} \lim_{m=\infty} s_{mn}$ *implies the existence of the other two; and all three are equal.*

Let us now assume that all the rows of the sequence converge, *i.e.* that $\lim_{n=\infty} s_{mn}$ exists for each value of m.

If $\lim_{n=\infty} s_{mn}$ be, for every value of m, numerically less than some fixed positive number, and if further, corresponding to any arbitrarily chosen positive number ϵ, an integer n_ϵ, dependent on ϵ, but *independent of* m, can be determined, such that $\left|s_{mn} - \lim_{n=\infty} s_{mn}\right| < \epsilon$, provided $n \geqq n_\epsilon$, and for every value of m, then the convergence of the rows is said to be *uniform*. A similar definition applies to the uniform convergence of the columns.

The rows of a double sequence may be uniformly convergent, and yet the columns need not converge.

From the theory of the limits of a function of two variables, the following theorem is easily deduced:—

In order that the double sequence $\{s_{mn}\}$, *of which the rows are known to converge, may converge, it is necessary and sufficient that the convergence of the rows be uniform, and that* $\lim_{m=\infty} \lim_{n=\infty} s_{mn}$ *shall exist.*

The double sequence may however be convergent, without the rows being convergent.

337. The theorems of § 233 and § 234 may be employed to obtain the necessary and sufficient conditions that the two limits

$$\lim_{m=\infty} \lim_{n=\infty} s_{mn}, \quad \lim_{n=\infty} \lim_{m=\infty} s_{mn}$$

may both exist, and have the same finite value.

We thus obtain the following theorems:—

In order that the repeated limits $\lim_{m=\infty} \lim_{n=\infty} s_{mn}$, $\lim_{n=\infty} \lim_{m=\infty} s_{mn}$ *may both exist and have the same finite value, it is necessary and sufficient,* (1) *that* $\overline{\lim}_{n=\infty} s_{mn} - \underline{\lim}_{n=\infty} s_{mn}$ *should have the limit zero, for* $m = \infty$, *and also that* $\overline{\lim}_{m=\infty} s_{mn} - \underline{\lim}_{m=\infty} s_{mn}$ *should have the limit zero, for* $n = \infty$; *and* (2) *that, corresponding to each fixed positive number* ϵ, *arbitrarily chosen, a positive integer* N *can be determined, such that for each value of* n *that is* $> N$, *a positive integer* M_n, *in general dependent on* n, *can be determined, such that, for this value of* n, s_{mn} *lies between* $\overline{\lim}_{n=\infty} s_{mn} + \epsilon$ *and* $\underline{\lim}_{n=\infty} s_{mn} - \epsilon$, *for all values of* m *that are* $> M_n$.

If M_n, *when found for* n, *is also applicable for all greater values of* n, *then the conditions, thus rendered more stringent, ensure that* $\lim_{m=\infty, n=\infty} s_{mn}$ *also exists, and is equal to the repeated limits.*

If the rows and the columns of the sequence* $\{s_{mn}\}$ *be both convergent,*

* This theorem was given by Bromwich, *Proc. Lond. Math. Soc.*, ser. 2, vol. I, p. 185; except that in the statement there given a redundant condition is contained, viz. that $\lim_{n=\infty, m=\infty} \lim s_{mn}$ must exist; this condition is however contained in the one stated in the theorem above. This arose from the fact that Bromwich deduced the theorem from a theorem corresponding to the alternative theorem given below, in which this condition is required.

i.e. if $\lim_{n=\infty} s_{mn}$, $\lim_{m=\infty} s_{mn}$ *both exist as definite numbers, then the necessary and sufficient conditions that the two repeated limits may both exist and have the same finite value is that, corresponding to each fixed positive number* ϵ, *arbitrarily chosen, a positive integer* N *can be determined, such that for each value of* n *that is* $> N$, *a positive integer* M_n, *in general dependent on* n, *can be determined, such that, for this value of* n, $\left| s_{mn} - \lim_{n=\infty} s_{mn} \right| < \epsilon$, *for all values of* m *that are* $> M_n$.

The following alternative set of conditions for the existence and equality of the repeated limits is obtained from the theorem in § 235:—

The necessary and sufficient conditions that $\lim_{m=\infty} \lim_{n=\infty} s_{mn} = \lim_{n=\infty} \lim_{m=\infty} s_{mn}$, *their value being finite, are* (1) *that* $\overline{\lim_{m=\infty}}\, s_{mn}$ *converge to a definite value* $\lim_{n=\infty} \lim_{m=\infty} s_{mn}$, *when* n *is indefinitely increased, and that* $\overline{\lim_{n=\infty}}\, s_{mn} - \underline{\lim_{n=\infty}}\, s_{mn}$ *converge to zero, as* m *is indefinitely increased;* and (2) *that, corresponding to each arbitrarily chosen positive number* ϵ, *and to an arbitrarily chosen integer* N, *a value* $N_1 (> N)$ *of* n *can be determined, and also an integer* M, *such that the condition that* s_{mN_1} *lies between* $\overline{\lim_{n=\infty}}\, s_{mn} + \epsilon$ *and* $\underline{\lim_{n=\infty}}\, s_{mn} - \epsilon$ *is satisfied for every value of* m *that is* $> M$.

In case $\lim_{m=\infty} s_{mn}$ *exists for every value of* m *except for a finite number of such values, the condition* (2) *is that* $\left| s_{mN_1} - \lim_{n=\infty} s_{mn} \right| < \epsilon$, *for every value of* m *that is* $> M$.

338. The preceding results may be applied to questions concerning the convergence of a double series

$$\begin{array}{l} a_{11} + a_{12} + a_{13} + \ldots + a_{1n} + \ldots \\ + a_{21} + a_{22} + a_{23} + \ldots + a_{2n} + \ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ + a_{m1} + a_{m2} + a_{m3} + \ldots + a_{mn} + \ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{array}$$

which has been defined in § 335, as a series of type ω^2. For this purpose we denote by s_{mn} the sum of the finite series

$$\begin{array}{l} a_{11} + a_{12} + a_{13} + \ldots + a_{1n} \\ + a_{21} + a_{22} + a_{23} + \ldots + a_{2n} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ + a_{m1} + a_{m2} + a_{m3} + \ldots + a_{mn}. \end{array}$$

The condition of convergence of the double series is then equivalent to that of the double sequence $\{s_{mn}\}$. If $\lim_{m=\infty, n=\infty} s_{mn}$ be $+\infty$ or $-\infty$, the double series is divergent; if $\lim_{m=\infty, n=\infty} s_{mn}$ does not exist, then the double series is said to oscillate.

In case the double series be convergent, it is not necessary that the single rows, or the single columns, should separately converge; but only a finite number of rows, or of columns, can diverge, or have infinite limits of indeterminacy; and the difference between the limits of indeterminacy of a row, or of a column, must be arbitrarily small, from and after some fixed row, or column.

If the double series be convergent, and if also every row be convergent, then the series of the sums of the rows must converge to the limits. A similar statement applies to the columns.

When all the rows, and all the columns, converge, and when the sum of these sums in each case converges, the double series may oscillate, and this is necessarily the case if the two limiting sums are different.

The series

$$a_{11} + a_{12} + a_{21} + a_{13} + a_{22} + a_{31} + \ldots + a_{1n} + a_{2(n-1)} + \ldots + a_{n1} + \ldots,$$

which has the type ω, is said to be the *diagonal* series corresponding to the double series. If this series converge, its sum is said to be the *diagonal sum* of the given series.

In accordance with a theorem in § 335, if all the terms of the double series be positive, the existence of the sum s ensures also that of the diagonal sum; and the converse is also the case. The convergence of the double series also ensures, in this case, that all the rows converge, and that the sum of their sums converges also to s. A similar statement holds for the columns; the sum of the series being independent of the type.

A convergent double series is said to be *absolutely convergent*, if the double series of which the terms are $|a_{mn}|$ be convergent*.

The theorem of § 333, that the absolute convergence of a series of any term implies the convergence of the series, and of all the series obtained by rearranging the terms in another type, shews that, if the given series be absolutely convergent, then the diagonal series converges to the sum of the double series; and also the sum taken by rows, or by columns, converges to the sum of the double series.

* It has been asserted by Jordan that there exist only absolutely convergent double series; see his *Cours d'Analyse*, vol. I, p. 302. This statement rests upon a defective definition of convergence.

Thus, for an absolutely convergent series, each one of the four equations

$$\sum_{m,n}^{\infty} a_{mn} = s, \quad \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} a_{mn} = s, \quad \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} a_{mn} = s,$$

$$\sum_{n=1}^{\infty}(a_{1n} + a_{2(n-1)} + \dots + a_{n1}) = s,$$

implies the other three.

A convergent double series which is not absolutely convergent can be replaced by a new series which diverges, and is such that each term a_{mn} occurs in a definite place in the new series, and that no terms occur in the new series which do not belong to the original one.

EXAMPLES.

1. Let $s_{mn} = (-1)^{m+n}(m^{-p} + n^{-q})$, where $p > 0$, $q > 0$. In this case $\lim_{n=\infty} s_{mn}$ and $\lim_{m=\infty} s_{mn}$ do not exist as definite numbers, but the three limits $\lim_{m=\infty,\, n=\infty} s_{mn}$, $\lim_{m=\infty}\lim_{n=\infty} s_{mn}$, $\lim_{n=\infty}\lim_{m=\infty} s_{mn}$ all exist, and are zero.

2. Let $s_{mn} = \dfrac{1}{1+(m-n)^2}$. In this case $\lim_{m=\infty,\, n=\infty} s_{mn}$ does not exist, but the two repeated limits both exist, and are zero. The same remarks apply to the case $s_{mn} = \dfrac{mn}{m^2+n^2}$.

3. Let $s_{mn} = \dfrac{(-1)^{m+n}}{2(a+1)}\left(\dfrac{1}{a^m} + \dfrac{1}{a^n}\right)$, where $a > 1$. In this case the double limit exists and is zero, but neither the rows nor the columns of the corresponding series are convergent, but are oscillating series; consequently the double series does not converge absolutely.

4. Let $a_{mn} = (-1)^{m+n}\left(\dfrac{1}{m+n-1} + \dfrac{1}{m+n}\right)$. In this case the single rows and the single columns converge, and the double series converges, but the sum of the diagonal series oscillates between $\log 2 + 1$ and $\log 2 - 1$.

FUNCTIONS REPRESENTED BY SERIES.

339. Let $$u_1(x),\ u_2(x),\ \dots\ u_n(x)\ \dots$$

be an unending sequence of functions, defined for a given domain of the variable x, which domain is most usually a continuous interval (a, b), but may be any given set of points G. The infinite series

$$u_1(x) + u_2(x) + \dots + u_n(x) + \dots$$

is taken to define a function $s(x)$, for the domain of the variable, in the following manner:—At any point $x = \alpha$, for which the series

$$u_1(\alpha) + u_2(\alpha) + \dots + u_n(\alpha) + \dots$$

converges, the limiting sum of the series is taken to be the value $s(a)$ of the function; if, at the point a, the series

$$u_1(a) + u_2(a) + \dots$$

diverges, the function $s(x)$ is undefined, but it is frequently convenient to say that the value of the function at that point is one of the improper numbers

$$+\infty, \text{ or } -\infty, \text{ or } \pm\infty,$$

according to the mode of divergence of the series. If, at $x = a$, the series is an oscillating one, the function $s(x)$ may be regarded as multiple-valued, and as having all the values to which a sequence

$$s_{n_1}(a), s_{n_2}(a), \dots$$

of the partial-sums may converge, $s(a)$ having thus the same limits of indeterminacy as the series itself.

If $$u_1(x) + u_2(x) + \dots + u_n(x)$$
be denoted by $s_n(x)$, the function $s(x)$ is definable as the limit of the sequence of functions

$$s_1(x), s_2(x), \dots s_n(x) \dots.$$

It will be observed that the term "limit" is here used in an extended sense, which covers the cases when, at a point a, the sequence of functional values is divergent or is oscillating. Stated in this form, the theory may be regarded as a theory of functions *defined as the limits of sequences of given functions*, the serial form being, in fact, only a particular mode of presentation. Thus

$$s_1(x), s_2(x), \dots s_n(x), \dots$$

may be a sequence of functions represented in any manner, for example by continued fractions, or by determinants; but, in whatever manner the $s_n(x)$ be represented, the limiting function $s(x)$ can always, of course, be exhibited in the form of the series

$$s_1(x) + [s_2(x) - s_1(x)] + [s_3(x) - s_2(x)] + \dots.$$

The function $s(x)$ may be termed the *sum-function* of the series.

UNIFORM CONVERGENCE OF SERIES.

340. If the series

$$u_1(x) + u_2(x) + \dots + u_n(x) + \dots$$

converge, for the point $x = a$, in the domain for which the functions $u(x)$ are defined, then, corresponding to each arbitrarily assigned positive number ϵ, an integer n can be found such that

$$|R_{n,1}(a)|, |R_{n,2}(a)|, \dots |R_{n,s}(a)| \dots$$

are all numerically less than ϵ, this being the condition of convergence of the series at the point a.

A similar statement holds as regards each point at which the series converges. The least value of n for which the condition stated is satisfied will in general depend upon the arbitrarily chosen number ϵ, and also upon the value of a; but it is important to consider the case in which n can be chosen, for each fixed ϵ, so as to be independent of a. Let it now be assumed that the series is everywhere convergent in a given domain.

If a value of n can be found, corresponding to each arbitrarily assigned positive number ϵ, such that, for all values of x which belong to a given domain,

$$|R_{n,1}(x)|, |R_{n,2}(x)|, \ldots |R_{n,s}(x)| \ldots$$

be all less than ϵ, then, if this value of n be independent of x, the series

$$u_1(x) + u_2(x) + u_3(x) + \ldots$$

is said to converge uniformly in the given domain of x.

If we denote by $\phi(\epsilon, x)$ the least value which n must have, for a fixed value of x belonging to the given domain, in order that

$$|R_{n,1}(x)|, |R_{n,2}(x)|, \ldots$$

may all be less than $\frac{1}{2}\epsilon$, the series is, in accordance with the above definition, uniformly convergent in the domain of x, provided that, corresponding to each fixed value of ϵ, the values of $\phi(\epsilon, x)$ for all values of x in the domain be all less than some fixed integer n_1; and this integer n_1 is such that all the numbers $|R_{n,s}(x)|$, for every value of x, are less than ϵ, for all values of n that are $\geqq n_1$. The definition which has been given of uniform convergence in a given domain includes the condition that the series converges at each point of the domain. If it be assumed that this is already known to be the case, the definition of uniform convergence may be stated as follows:—

If the series $\quad u_1(x) + u_2(x) + \ldots + u_n(x) + \ldots$

converge for each value of x in a given domain to the value $s(x)$, then the series is said to converge uniformly in the domain, provided that, corresponding to each arbitrarily chosen positive ϵ, a number n, independent of x, can be found, such that all the remainders

$$|s(x) - s_n(x)|, \; |s(x) - s_{n+1}(x)|, \ldots |s(x) - s_{n+m}(x)| \ldots,$$

for every value of x, are less than ϵ.

341. A mode of convergence of a series in a given domain, less stringent in character than that of uniform convergence, has been considered by Dini and by other writers. This mode of convergence has been termed by Dini "simple-uniform convergence," and is defined by him* as follows:—

* See *Grundlagen*, by Lüroth and Schepp, p. 137.

The series $u_1(x) + u_2(x) + \dots + u_n(x) + \dots$

which converges at each point of a given domain to the value $s(x)$, ***is said to be simply-uniformly convergent in the domain, if, corresponding to each arbitrarily chosen positive number*** ϵ, *and to each integer* m', ***only one or several integers*** m, *not less than* m', *exist, which are such that, for all values of* x *in the domain, the* $|R_m(x)|$ *are* $< \epsilon$.

The condition of simple-uniform convergence is less stringent than that of uniform convergence, in that, in the latter case, all the remainders after a certain one are numerically less than ϵ, whereas in the former case one or several, but not all the remainders, need be numerically less than ϵ.

As regards the above definition, it may be remarked that, if there be, for each ϵ, one value of m which satisfies the prescribed condition, there must be an infinite number of such values; because we have only to ascribe to m' a series of values which increase indefinitely, and for each of these exists a corresponding value of m. Any one of an infinite set of values of m may thus be taken to correspond to one value of m'. Moreover, the definition can be reduced to a simpler form, thus:—Let us first suppose that there exists no value of n such that $R_n(x) = 0$ for every value of x in the domain; it will then be shewn that, if, corresponding to each ϵ, one value of n can be found such that $|R_n(x)| < \epsilon$, independently of x, then there must be an indefinitely great number of such values of n. Let us denote by $\bar{R}_s$, the upper limit of $|R_s(x)|$ for the whole domain of x; $\bar{R}_s$ may have a definite value, or it may be indefinitely great. If $|R_n(x)| < \epsilon$ for every value of x, we have $\bar{R}_n \leqq \epsilon$; let us therefore take a positive number ϵ_1 less than $\bar{R}_n$, and also less than the least of the numbers

$$\bar{R}_1, \bar{R}_2, \dots \bar{R}_{n-1},$$

then, by hypothesis, a number n_1 can be found such that, for all values of x in the domain, $|R_{n_1}(x)| < \epsilon_1$. This number n_1 cannot be one of the numbers $1, 2, 3, \dots n$; for it is always possible to find a value of x for which $|R_s(x)|$ is arbitrarily near its upper limit $\bar{R}_s$, and is thus $> \epsilon_1$; hence a number $n_1, > n$, has been shewn to exist, such that, for all the values of x, $|R_{n_1}(x)| < \epsilon$. Similarly it may be shewn that a number $n_2, > n_1$, exists which has the same property; and thus there is an indefinitely great set of values of n such that $|R_n(x)| < \epsilon$. If there be an indefinitely great number of values of n such that $R_n(x) = 0$, for every value of x in the domain, Dini's definition of uniform convergence is satisfied. In the case in which there are a finite number of such values of n, it will be sufficient, in order to ensure simple-uniform convergence, that, for each ϵ one value of n shall exist, such that $|R_n(x)| < \epsilon$, and also such that $R_n(x)$ is not zero for every value of x; in this case the above reasoning is applicable, provided ϵ_1 be taken less than $\bar{R}_n$, and also less than all those of the numbers $\bar{R}_1, \bar{R}_2, \dots \bar{R}_{n-1}$, which do not vanish. The definition of simple-uniform convergence may now be stated as follows:—

A series which converges for every value of x in a given domain is said to converge simply-uniformly either, (1) if there be at most a finite number of values of n such that $R_n(x)$ vanishes for every value of x, and if, corresponding to each arbitrarily chosen positive number ϵ, a number n can be found such that, independently of x, $|R_n(x)| < \epsilon$, whilst $R_n(x)$ does not vanish for every value of x, or (2) if there be an indefinitely great number of values of n for which $R_n(x)$ vanishes for every value of x.

A series which is uniformly convergent is also simply-uniformly convergent; but the converse does not hold.

If the series be simply-uniformly convergent, but be not uniformly convergent, there must, corresponding to each sufficiently small ϵ, be an indefinitely great number of values of n for which the condition

$$|R_n(x)| < \epsilon,$$

for all values of x, is not satisfied; for if there were only a finite number of such values, n could be taken greater than the greatest of these, and thus the condition for uniform convergence would be satisfied, which would be contrary to hypothesis.

If all the terms $u_n(x)$ of a series be positive for every value of x in the domain of the variable, then, if the series $\Sigma\, u(x)$ be simply-uniformly convergent, it is also necessarily uniformly convergent. For the condition of simple-uniform convergence ensures that, corresponding to an arbitrarily chosen ϵ, n can be found such that the sequence

$$(R_{n,1}, R_{n,2}, R_{n,3} \ldots)$$

converges for every value of x to a value which is less than ϵ; and, since the terms of the series are all positive, each element of this sequence is less than, or equal to, the next one; and therefore

$$R_{n,1}, R_{n,2}, R_{n,3} \ldots \text{ are all } < \epsilon.$$

It follows that $R_{n+m,s}$, which equals $R_{n,s+m} - R_{n,m}$, is also $< \epsilon$, for all values of m and s; and thus that $R_{n+m} < \epsilon$; hence the series converges uniformly.

It may easily be shewn that, if the two series $\Sigma\, u(x)$, $\Sigma\, |u(x)|$, be both simply-uniformly convergent, then $\Sigma\, u(x)$ is uniformly convergent.

Let $\epsilon_1, \epsilon_2, \ldots$ be a sequence of diminishing positive numbers which converges to zero. If the series $\Sigma\, u(x)$ be simply-uniformly convergent, a number n_1 can be found such that $|R_{n_1}(x)| < \epsilon_1$, for all values of x; a number $n_2, > n_1$, can then be found such that $|R_{n_2}(x)| < \epsilon_2$; then n_3, such that $|R_{n_3}(x)| < \epsilon_3$; and so on. If now the first n_1 terms be amalgamated into one, then those after the first n_1 up to and including $u_{n_2}(x)$, and so on, the series may be written in the form

$$s_{n_1}(x) + [s_{n_2}(x) - s_{n_1}(x)] + [s_{n_3}(x) - s_{n_2}(x)] + \ldots;$$

and in this form the series is uniformly convergent. It thus appears that *a simply-uniformly convergent series can be changed into one which is uniformly convergent, by bracketing the terms suitably, in accordance with a norm, and taking each bracket to constitute a term in the new series.*

Conversely, *a uniformly convergent series may be replaced by one which is only simply-uniformly convergent.*

If each term $u_n(x)$ of a uniformly convergent series be replaced, in accordance with some norm, by the sum of r_n functions, such that

$$u_n(x) = U_{n,1}(x) + U_{n,2}(x) + \dots + U_{n,r_n}(x),$$

then the new series

$$U_{1,1}(x) + U_{1,2}(x) + \dots + U_{1,r_1}(x) + U_{2,1}(x) + \dots$$

is not necessarily convergent, but may oscillate. If, however, the functions U be so chosen that the series converges in the whole domain of x, then the series converges at least simply-uniformly. It thus appears* that the distinction between uniform convergence and simple-uniform convergence is an unessential one.

NON-UNIFORM CONVERGENCE.

342. If we denote by $\psi(\epsilon, x)$ the least value which n can have, such that

$$|R_n(x)|,\ |R_{n+1}(x)|,\ |R_{n+2}(x)| \dots$$

may all be $< \epsilon$, where the series $\Sigma u(x)$ converges, for every value of x in a given domain, to the value $s(x)$, then the condition of convergence ensures that for any fixed value of x, $\psi(\epsilon, x)$ has a definite finite value, for each value of ϵ, which however may increase indefinitely as ϵ is indefinitely diminished. Taking a fixed value of ϵ, sufficiently small, it may happen that $\psi(\epsilon, x)$ has no finite upper limit for all values of x in the domain; and this will happen in case the convergence of the series be non-uniform. The function $\{\psi(\epsilon, x)\}^{-1}$ has, in this case, zero for its lower limit; and therefore in accordance with the theorem of § 171, there must be at least one point x, such that, in an arbitrarily small neighbourhood of it, zero is the lower limit of $\{\psi(\epsilon, x)\}^{-1}$. There may be a finite, or an infinite, set of such points; and, in an arbitrarily small neighbourhood of any one point of this set, $\psi(\epsilon, x)$ has no upper limit, and thus has values greater than any arbitrarily chosen number A. Nevertheless $\psi(\epsilon, x)$ has a definite finite functional value at each point of the set, provided that such a point belong to the domain of the variable; for otherwise the series would not converge at such a point.

* See Arzelà, *Bologna Rendiconti*, 1899; also Hobson, *Proc. Lond. Math. Soc.*, ser. 2, vol. I, p. 376.

A point, in the arbitrarily small neighbourhood of which $\psi(\epsilon, x)$ has no upper limit, provided ϵ be sufficiently small, is said to be such that the series is non-uniformly convergent in its neighbourhood.

Frequently, for shortness, such a point is said to be a point *at* which the series is non-uniformly convergent, or to be *a point of non-uniform convergence.*

It has been shewn that such points exist whenever the series is non-uniformly convergent in the domain of the variable; and, in the case in which the domain is a closed set, which case is alone of importance, these points themselves all belong to the domain of the variable.

When there are only a finite number of points of non-uniform convergence, the series is frequently said to be *in general* uniformly convergent. It becomes in this case, uniformly convergent, if arbitrarily small intervals containing these points be removed from the domain of the variable.

If $x = \alpha$, be a point of non-uniform convergence, a sequence

$$\alpha_1, \alpha_2, \ldots \alpha_n, \ldots$$

of values of x in the closed domain can be found which converges to the value α, and is such that the numbers

$$\psi(\epsilon, \alpha_1), \psi(\epsilon, \alpha_2), \ldots \psi(\epsilon, \alpha_n) \ldots$$

form a sequence with no upper limit, where ϵ has a fixed value chosen sufficiently small. Thus one of the limits

$$\psi(\epsilon, \overline{\alpha+0}), \psi(\epsilon, \overline{\alpha-0})$$

is infinite, or both are so, although $\psi(\epsilon, \alpha)$ must be itself finite. Therefore α is a point of infinite discontinuity of the function $\psi(\epsilon, x)$.

If one, but not both, of the limits

$$\psi(\epsilon, \overline{\alpha+0}), \psi(\epsilon, \overline{\alpha-0})$$

be infinite, the point is said to be one of *non-uniform continuity on the right* or *on the left*, as the case may be.

THE CONTINUITY OF THE SUM-FUNCTION.

343. Let us suppose that the domain of x is either the interval (a, b), or else a perfect set of points in that interval, and further that the functions

$$u_1(x), u_2(x), u_3(x), \ldots$$

are continuous throughout the domain. It will then be shewn that:—

If the series $\Sigma u(x)$ converge at least simply-uniformly in the domain of the variable, the sum-function $s(x)$ is everywhere continuous.

Let α be any point in the domain of x, and $\alpha+\delta$ another such point on the right of α; then

$$s(\alpha)=s_n(\alpha)+R_n(\alpha),$$

$$s(\alpha+\delta)=s_n(\alpha+\delta)+R_n(\alpha+\delta);$$

thus

$$s(\alpha+\delta)-s(\alpha)=[s_n(\alpha+\delta)-s_n(\alpha)]+[R_n(\alpha+\delta)-R_n(\alpha)].$$

Since the series converges simply-uniformly, a value of n can be found, corresponding to any arbitrarily small ϵ, such that

$$|R_n(\alpha)|,\ |R_n(\alpha+\delta)| \text{ are each } <\tfrac{1}{4}\epsilon,$$

for all positive values of δ such that $\alpha+\delta$ belongs to the domain of x. Suppose n to have this value; then, since $s_n(x)$ is continuous, a value δ_1 of δ can be determined such that

$$|s_n(\alpha+\delta)-s_n(\alpha)|<\tfrac{1}{2}\epsilon, \text{ if } 0<\delta\leqq\delta_1.$$

It follows that $|s(\alpha+\delta)-s(\alpha)|<\epsilon$, provided $0<\delta\leqq\delta_1$;

and, as ϵ has been arbitrarily chosen, the condition of continuity of $s(x)$ at α, on the right, is satisfied. In a similar manner it may be shewn that $s(x)$ is continuous at α on the left.

A fortiori, the condition that the series converge uniformly is sufficient to secure that the sum-function may be continuous.

The above proof also suffices to establish the following more general theorem :—

If the functions $u_n(x)$ be all continuous at the point $x=\alpha$, but not necessarily elsewhere, the condition of simple-uniform convergence of the series in an interval containing the point α in its interior is sufficient to ensure that $s(x)$ is continuous at the point α.

344. If the function $s(x)$ be discontinuous at α, say on the right, then δ_1 cannot be chosen so that, for

$$0<\delta\leqq\delta_1,\ |s(\alpha+\delta)-s(\alpha)|<\epsilon,$$

provided ϵ be chosen sufficiently small; hence, in this case, it is impossible to choose n such that

$$|R_n(\alpha+\delta)-R_n(\alpha)|<\tfrac{1}{2}\epsilon,$$

for all the values of δ concerned, s_n being a continuous function for the domain; and it follows that it is impossible to choose n such that

$$|R_n(\alpha+\delta)|<\tfrac{1}{4}\epsilon,$$

for all values of δ such that $0<\delta\leqq\delta_1$.

Therefore, in this case, the series converges neither uniformly nor simply-uniformly, and the point α is a point of non-uniform convergence.

It has long been known that the sum of a series of which all the terms are continuous is not necessarily itself continuous. The important discovery that such a discontinuity is due to the non-uniform convergence of the series was made independently by Stokes* and by Seidel†. It was not until a later time that, under the influence of Weierstrass, the great importance of the notion of uniform convergence in the Theory of Functions was fully recognized.

The question whether non-uniform convergence necessarily implies discontinuity in the sum-function remained for many years an open one. It was decided in the negative sense when Darboux and Du Bois Reymond constructed examples of cases in which the series are non-uniformly convergent, and yet nevertheless have continuous sum-functions.

TESTS OF UNIFORM CONVERGENCE.

345. In certain cases it can be easily established that a series is uniformly convergent. This can frequently be done by applying the following theorem :—

$\Sigma u_n(x)$ denoting a series of functions such that $|u_n(x)|$ has, for each value of n, an upper limit $\bar{u}_n$ for the whole domain, if the series

$$\bar{u}_1 + \bar{u}_2 + \dots + \bar{u}_n + \dots$$

be convergent, then the given series is itself uniformly convergent, and is absolutely convergent for each value of x.

The remainder $$\bar{u}_{n+1} + \bar{u}_{n+2} + \dots$$

of the series $\Sigma\bar{u}$, is greater than, or equal to, the remainder

$$|u_{n+1}(x)| + |u_{n+2}(x)| + \dots$$

of the series $\Sigma|u_n(x)|$. If n be so chosen that the former remainder be $< \epsilon$, the latter remainder is also $< \epsilon$, for every value of x; and the convergency condition of $\Sigma\bar{u}_n$ states that, corresponding to each ϵ, a number n_1 exists, such that all the remainders, of index $\geqq n_1$, are $< \epsilon$; therefore the same holds as regards the series $\Sigma|u_n(x)|$. Hence this latter series is uniformly convergent; and since no remainder of $\Sigma u_n(x)$ can exceed numerically the corresponding one of $\Sigma|u_n(x)|$, it follows that $\Sigma u(x)$ is uniformly convergent, and converges absolutely for each value of x in the given domain.

346. *If all the terms of the series $u_1(x) + u_2(x) + \dots + u_n(x) + \dots$ be positive or zero, for all values of x in (a, b), and the series converge uniformly in that interval, then the series obtained by rearranging the order of the terms, in accordance with some norm, is also uniformly convergent in the interval.*

That the new series obtained by rearranging the order of the terms

* "On the critical values of the sums of periodic series," *Math. and Physical Papers*, vol. I, p. 236.

† "Note über eine Eigenschaft der Reihen," *Abhd. d. Münch. Akad.* vol. VII. On the history of this discovery, see Reiff's *Geschichte der unendlichen Reihen*, p. 207.

converges to the sum of the original series, everywhere in (a, b), has been proved in § 332. An integer n' exists, such that the first n terms of the given series all occur amongst the first n' terms of the new series; it follows that $R_n(x) \leqq R'_{n'}(x)$, where $R_n(x)$, $R'_{n'}(x)$ denote the remainders after n and n' terms respectively, in the original series and in the new series. If n be so chosen that $R_n(x) < \epsilon$, for all values of x in (a, b), we have also $R'_{n'}(x) < \epsilon$ for all values of x; therefore the new series is also uniformly convergent.

If the series $|u_1(x)| + |u_2(x)| + \ldots + |u_n(x)| + \ldots$ *converge uniformly in* (a, b), *then the series* $u_1(x) + u_2(x) + \ldots + u_n(x) + \ldots$ *converges uniformly in* (a, b); *also any series obtained by rearranging the order of the terms of the latter series, in accordance with any norm, is uniformly convergent.*

The second series is necessarily convergent everywhere in (a, b); also its remainder after n terms cannot exceed, in absolute value, the remainder after n terms of the first series. It follows that, if n be so chosen that the remainder of the first series after n terms is less than ϵ, for every value of x, the absolute value of the remainder of the second series satisfies the same condition. Therefore the second series is uniformly convergent. Since, from the last theorem, a rearrangement of the order of terms of the first series does not affect its uniform convergence, it follows that a corresponding rearrangement of the terms of the second series does not affect its uniform convergence.

The converse of this theorem also holds, and may be stated as follows:—

If the series* $u_1(x) + u_2(x) + \ldots$ *be uniformly convergent in* (a, b), *and if all the series obtained by systematic rearrangement of the terms of the series be also uniformly convergent, then the series* $|u_1(x)| + |u_2(x)| + \ldots$ *is uniformly convergent in the same interval.*

347. The following theorem† is sometimes useful:—

If the terms of the series $u_1(x) + u_2(x) + \ldots$ *be continuous in* (a, b), *and never negative, and if the series converge to a continuous sum-function* $s(x)$, *then the series converges uniformly in* (a, b).

To prove this theorem, let x_1 be any point in (a, b), then

$$s(x) - s(x_1) = \{s_n(x) - s_n(x_1)\} + \{R_n(x) - R_n(x_1)\}.$$

For the fixed point x_1, and corresponding to any fixed positive number ϵ, an integer n can be so chosen that $R_n(x_1) < \frac{1}{3}\epsilon$. This value of n being fixed, an interval $(x_1 - \delta, x_1 + \delta)$ can be so determined, that, if x be in this interval, both $|s(x) - s(x_1)|$ and $|s_n(x) - s_n(x_1)|$ are $< \frac{1}{3}\epsilon$; this follows from the con-

* This theorem has been proved by G. D. Birkhoff, *Annals of Math.*, ser. 2, vol. VI, 1905, p. 90.

† See Dini's *Grundlagen*, p. 148.

tinuity of $s(x)$ and $s_n(x)$ at x_1. We now see that, throughout the interval $(x_1 - \delta, x_1 + \delta)$, the condition $R_n(x) < \epsilon$ is satisfied; and since the terms of the given series are never negative, it follows that $R_{n'}(x) < \epsilon$, for every value of n' that is $\geqq n$, and every value of x in $(x_1 - \delta, x_1 + \delta)$. It has therefore been proved that x_1 is a point of uniform continuity of the series; and since x_1 is any point whatever in (a, b), the convergence of the series is uniform in (a, b).

If a sequence $s_1(x), s_2(x), \ldots s_n(x), \ldots$ be such that, for every value of x in an interval (a, b), one of the sets of conditions $s_n(x) \geqq s_{n+1}(x)$ for every value of n, or $s_n(x) \leqq s_{n+1}(x)$ be satisfied, then the sequence is said to be *monotone* in (a, b).

The above theorem may be stated in the following form:—

A sequence of continuous functions $\{s_n(x)\}$ which is monotone in a given interval, and which converges to a continuous function $s(x)$, converges uniformly to $s(x)$.

348. *If* $u_1(x), u_2(x), \ldots u_n(x), \ldots$ be defined for the interval (a, b), and be limited in that interval, and positive for all the values of x, and if further $u_n(x) \geqq u_{n+1}(x)$, for every value of n and x; then, if Σa_n be any convergent series, the series $\Sigma a_n u_n(x)$ converges uniformly in the interval (a, b). Moreover, if Σa_n do not converge, but oscillate between finite limits of indeterminacy, then, provided the additional conditions that the functions $u_n(x)$ be all continuous, and that $\lim_{n=\infty} u_n(x) = 0$ for each value of x, be satisfied, the series $\Sigma a_n u_n(x)$ is uniformly convergent in the interval (a, b), and its sum is consequently continuous.*

In case the series Σa_n be convergent, the partial remainder $R_{n,m}$ of the series $\Sigma a_n u_n(x)$ being

$$(a_{n+1} + a_{n+2} + \ldots + a_{n+m})\, u_{n+m+1}(x) + \sum_{r=1}^{r=m} (a_{n+1} + a_{n+2} + \ldots + a_{n+r})\{u_{n+r}(x) - u_{n+r+1}(x)\},$$

we see that, by choosing n so great that all the partial remainders of the series Σa_n after the nth term are numerically less than the arbitrarily chosen number ϵ, the condition

$$|R_{n,m}| < \epsilon u_{n+1}(x)$$

is satisfied; and therefore, for every value of x in (a, b), we have $|R_{n,m}| < \epsilon . U$, where U denotes the upper limit of $u_1(x)$ in (a, b). Since ϵU is arbitrarily small, it has thus been shewn that the condition of uniform convergence of $\Sigma a_n u_n(x)$ is satisfied.

It is easily seen that this part of the above theorem also holds when the terms of the series Σa_n are functions of x, provided Σa_n converges uniformly

* See Hardy, *Proc. Lond. Math. Soc.*, ser. 2, vol. IV, pp. 250, 251.

in (a, b). When the series Σa_n oscillates between finite limits, K can be determined such that $|a_{n+1} + a_{n+2} + \dots + a_{n+r}| < K$, for all values of n and r. Also, since the sequence $u_1(x), u_2(x), u_3(x), \dots u_n(x), \dots$ is by hypothesis monotone, and converges to the continuous limit zero in the interval (a, b), it follows from the theorem of § 347, that the sequence converges uniformly to the limit zero. We can consequently choose n so that $u_{n+r}(x) < \epsilon$, for every value of r and x; therefore $|R_{n,m}(x)| < 3\epsilon K$, and since $3\epsilon K$ is arbitrarily small, it follows that the series is uniformly convergent.

EXAMPLES.

1. Let* $u_{2n-1}(x) = \dfrac{x}{nx^2 + (1-nx)^2}$, $\quad u_{2n}(x) = \dfrac{-x}{(n+1)x^2 + \{1-(n+1)x\}^2}$.

In this case, the series converges for all values of x, and

$$s(x) = \frac{x}{x^2 + (1-x)^2}; \qquad R_{2n-1}(x) = 0, \qquad R_{2n-2}(x) = u_{2n-1}(x).$$

In an interval (α, β), which contains the point $x=0$, the series converges simply-uniformly, but it does not converge uniformly, since $R_{2n-2}\left(\frac{1}{n}\right) = 1$, however great n may be.

2. Let† $s_n(x) = \dfrac{n^2 x}{1 + n^3 x^2}$, $s(x) = 0$, for $0 \leqq x \leqq 1$. This series converges non-uniformly

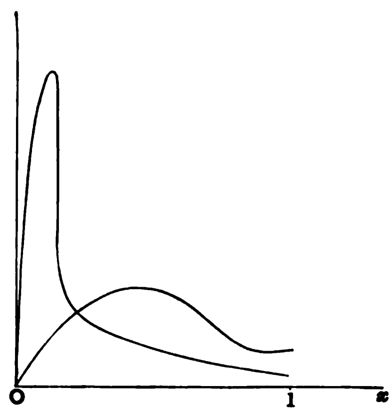


FIG. 1.

in the neighbourhood of the point $x=0$. The approximation curves $y = s_n(x)$ have peaks of height $\frac{1}{2}n^{\frac{1}{2}}$, which increase indefinitely in height as n is increased. At the same time, the point $\dfrac{1}{n^{\frac{3}{2}}}$, at which the ordinate is a maximum, continually approaches the point 0; and thus, for any value of x which is >0, n may be taken so great that $s_n(x)$ is arbitrarily small. At the point $x=0$, we have $s_n(x)=0$, for every n.

* Tannery, *Théorie des fonctions*, p. 134.

† Osgood, *Amer. Journal of Math.*, vol. XIX, p. 156; also G. Cantor, *Math. Annalen*, vol. XVI, p. 269.

3. Let $s_n(x)=\frac{nx}{1+n^2x^2}$, $s(x)=0$, $0\leq x\leq 1$. The curves $y=s_n(x)$ have peaks all of the same height $\frac{1}{2}$ at the points $x=\frac{1}{n}$. As in the last example the point $x=\frac{1}{n}$, below the peak, continually approaches the origin as n is increased. The convergence is non-uniform in the neighbourhood of $x=0$.

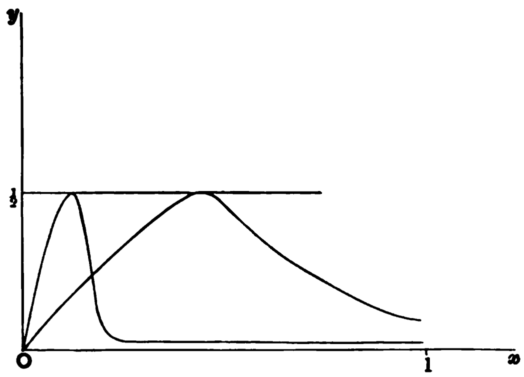


FIG. 2.

Let $\phi_k(x)=\frac{n\sin^2 k\pi x}{1+n^2\sin^4 k\pi x}$; $s_n(x)=\phi_{1!}(x)+\frac{1}{2!}\phi_{2!}(x)+\frac{1}{3!}\phi_{3!}(x)+\ldots=\sum_{1}^{\infty}\frac{1}{i!}\phi_{i!}(x)$.

The series which defines $s_n(x)$ converges uniformly, and thus $s_n(x)$ is a continuous function of x. In the neighbourhood of any rational point $x=p/q$, the curve $y=s_n(x)$ has peaks arising from the term $\frac{1}{k!}\phi_{k!}(x)$, where k is the smallest integer such that $k!$ is divisible by q. The series converges to the limit $s(x)=0$, non-uniformly in any interval whatever (a, b), taken in the interval $(0, 1)$.

4. Let* $u_{2n-1}(x)=x^{n+1}$, $u_{2n}(x)=-x^{n+1}\left\{1-\frac{1}{(n+1)!}\right\}$, where $0\leq x<1$, and

$$u_n(1)=\frac{1}{n^2}.$$

The series $\Sigma u(x)$ is simply-uniformly convergent in $(0, 1)$, but it is not uniformly convergent.

5. Let† $u_n(x)=x^n(1-x)$, $0\leq x\leq 1$. In this case $s(x)=x$, for $0\leq x<1$; but $s(x)=0$, for $x=1$; and the series converges non-uniformly in the neighbourhood of the point $x=1$.

6. Let† $u_n(x)=x^n(1-x^n)$. If $|x|<1$, we find $s(x)=\frac{x}{1-x^2}$; also $s(1)=0$; whereas $\lim_{x=1} s(x)$ is indefinitely great. The series converges non-uniformly in the neighbourhood of the point 1, and its sum-function has an infinite discontinuity at that point.

7. Let‡ $u_n(x)=-2(n-1)^2xe^{-(n-1)^2x^2}+2n^2xe^{-n^2x^2}$. Here $s(x)=0$ for every value of x; $R_n(x)=-2n^2xe^{-n^2x^2}$; and at $x=\frac{1}{n}$, $R_n\left(\frac{1}{n}\right)=-\frac{2n}{e}$. The series converges non-uniformly in

* Volterra, *Gior. di Mat.*, vol. XIX, p. 79.

† Arzelà, *Memorie di Bologna*, ser. 5, vol. VIII, p. 189.

‡ Darboux, *Ann. de l'école normale supérieure*, vol. V, "Sur les fonctions discontinues"

the neighbourhood of $x=0$, since arbitrarily large values of the $|R_n(x)|$ exist in such neighbourhood; but the sum-function is continuous at $x=0$.

8. Let* $$s_n(x)=\phi_n(x)+\frac{1}{2!}\phi_n(2!\,x)+\ldots+\frac{1}{k!}\phi_n(k!\,x)+\ldots$$

where $\phi_n(x)=\sqrt{2e}\,.\,n\sin^2\pi x\,.\,e^{-n^2\sin^4\pi x}$. The series which defines $s_n(x)$ converges uniformly, since $|\phi_n(k!\,x)|\leq 1$; and thus $s_n(x)$ is a continuous function of x. The sum-function $s(x)$ is also a continuous function of x; but the convergence of the functions $s_n(x)$ to $s(x)$ is non-uniform in every sub-interval of the interval (0, 1).

9. Consider† the series

$$\frac{1+5x}{2(1+x)}+\ldots+\frac{x(x+2)n^2+x(4-x)n+1-x}{n(n+1)\{(n-1)x+1\}(nx+1)}+\ldots.$$

Here $u_n(x)=\left[\frac{1}{n}+\frac{2}{(n-1)x+1}\right]-\left[\frac{1}{n+1}+\frac{2}{nx+1}\right]$; thus $s(x)=3$, unless $x=0$, when $s(0)=1$; and the sum-function is therefore discontinuous at the point 0.

Since $R_n(x)=\frac{1}{n+1}+\frac{2}{nx+1}$, we find on equating this to ϵ, and solving for n,

$$n=\{x+2-\epsilon(x+1)+\sqrt{[\{x+2-\epsilon(x+1)\}^2+4\epsilon x(3-x)]}\}/2\epsilon x;$$

thus, for a fixed ϵ, the value of n increases indefinitely as x approaches the value 0.

10. The series‡

$$x^2-x^2+\frac{x^2}{1+x^2}-\frac{x^2}{1+x^2}+\frac{x^2}{(1+x^2)^2}-\frac{x^2}{(1+x^2)^2}+\frac{x^2}{(1+x^2)^3}-\ldots$$

is uniformly and absolutely convergent in any interval $(-A,\ B)$. For $s_{2n}(x)=0$, $s_{2n+1}(x)=\frac{x^2}{(1+x^2)^n}$, and hence $s_{2n+1}(x)<\frac{1}{n}$; therefore the series converges uniformly to the sum zero. The series

$$x^2-x^2+\frac{x^2}{1+x^2}+\frac{x^2}{(1+x^2)^2}-\frac{x^2}{(1+x^2)^2}+\frac{x^2}{(1+x^2)^3}+\frac{x^2}{(1+x^2)^4}-\frac{x^2}{(1+x^2)^3}+\ldots,$$

obtained by rearranging the terms of the given series, is however non-uniformly convergent in $(-A,\ B)$. For $s_{3n-1}(x)=\frac{(1+x^2)^{n-1}-1}{(1+x^2)^{2n-2}}$; and for $x=\pm(2^{\frac{1}{n-1}}-1)^{\frac{1}{2}}$,

$$s_{3n-1}(x)=\tfrac{1}{4}.$$

The given series does not satisfy the condition stated in the theorem of § 346, that the series whose terms are the absolute values of those of the given series should be uniformly convergent. For the series

$$x^2+x^2+\frac{x^2}{1+x^2}+\frac{x^2}{1+x^2}+\frac{x^2}{(1+x^2)^2}+\ldots$$

has its sum discontinuous at the point $x=0$, and therefore does not converge uniformly in an interval $(-A,\ B)$.

* Osgood, "A geometrical method for the treatment of uniform convergence," *Bulletin of the American Math. Soc.*, 1896.

† Stokes, *Math. and Phys. Papers*, vol. I.

‡ Bôcher, *Annals of Math.* ser. 2, vol. IV, 1904, p. 159.

RELATION OF THE THEORY WITH THAT OF FUNCTIONS OF TWO VARIABLES.

349. If the functions $s_1(x), s_2(x), \ldots s_n(x), \ldots$ converge for all values of x in a given domain to the value $s(x)$, the functions $s_n(x)$, and

$$R_n(x) = s(x) - s_n(x),$$

may be regarded as functions of two variables x, y, where $n = 1/y$, and may be written $s(x, y)$, $R(x, y)$. These functions have been defined only for values of y which are the reciprocals of positive integers; it is however frequently convenient to assume that, for a value of y between two values y_m, y_{m+1}, which correspond to consecutive integral values $m, m+1$ of n, the functions are defined by

$$s(x, y) = \frac{y - y_m}{y_{m+1} - y_m}\, s(x, y_{m+1}) + \frac{y_{m+1} - y}{y_{m+1} - y_m}\, s(x, y_m),$$

$$R(x, y) = \frac{y - y_m}{y_{m+1} - y_m}\, R(x, y_{m+1}) + \frac{y_{m+1} - y}{y_{m+1} - y_m}\, R(x, y_m),$$

so that $s(x, y)$, $R(x, y)$ are continuous linear functions of y in the interval (y_m, y_{m+1}).

If we further assume

$$R(x, 0) = 0, \quad s(x, 0) = s(x),$$

the functions $s(x, y)$, $R(x, y)$ are defined for all values of x in the domain of x, and for all values of y in the interval $(0, 1)$, the ends included. These functions are everywhere continuous with respect to the variable y. That this is the case for $y = 0$, follows from the condition of convergence

$$\lim_{y=0} s(x, y) = s(x) = s(x, 0),$$

$$\lim_{y=0} R(x, y) = \quad 0 \quad = R(x, 0).$$

The functions $s(x, y)$, $R(x, y)$ may be termed* *the transformed sum-function*, and *the transformed remainder-function* respectively. The study of the properties of series or sequences of functions of a variable may be thus reduced to the study of the properties of functions of two variables, and this is frequently a very convenient procedure.

The function $R(x, y)$ is continuous with respect to x in the domain of x, upon the line $y = 0$, since it is everywhere zero; but it is not necessarily continuous with respect to (x, y).

* Hobson, "On non-uniform convergence and the integration of series," *Proc. Lond. Math. Soc.*, vol. XXXIV, p. 247.

Let P be a limiting point of the domain of x, on the x-axis; describe a semi-circle qpq' of radius ρ, with P as centre. The upper limit of $|R(x, y)|$ in this semi-circle will have a value $\beta(\rho)$ which is a function of ρ, and which has a limiting value β_P, when ρ is indefinitely diminished. It may happen that β_P is indefinitely great. If β_P be zero, P is a point of continuity of $R(x, y)$ with respect to (x, y); but if β_P be not zero, P is a point of discontinuity.

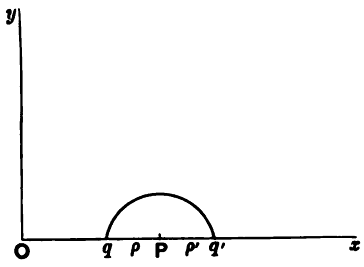


FIG. 3.

It is easily seen that a point P of discontinuity of $R(x, y)$ is a point of non-uniform convergence of the sequence $\{s_n(x)\}$.

At a point of uniform convergence, corresponding to an arbitrarily assigned number ϵ, in a sufficiently small neighbourhood NM, a value y_ϵ of y can be found, such that for $y \leqq y_\epsilon$, we have $|R(x, y)| < \epsilon$; and this will be the case for all points of the domain of x in the rectangle $MNSR$. Within this rectangle semi-circles with centre P can be described in which $|R(x, y)| < \epsilon$, for all points within the semi-circle, and thus the value of β_P is zero; and therefore P is a point of continuity of the transformed remainder-function.

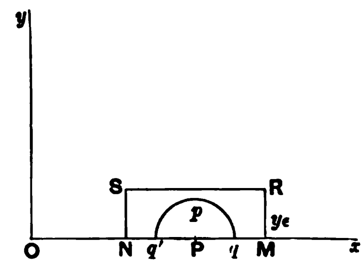


FIG. 4.

If at P, the number β_P be not zero, the convergence is non-uniform in the neighbourhood of the point P; and the number β_P, which is the saltus at P of $|R(x, y)|$, may be called *the* measure of non-uniform convergence at P*. The number β_P may be regarded as existent at every point which is a limiting point of the domain of x, although there may be points at which it has the improper value $+\infty$; and at a point of uniform convergence it has the value zero. If β is regarded as a function of the point P, it may be termed *the convergence-function**.

If the semi-circle used in defining β_P be divided into two quadrants by means of the radius Pp, the upper limits of $|R(x, y)|$ in the quadrants Ppq, Ppq' may be considered separately. When ρ has the limit zero, these upper

* The term "Grad der ungleichmässigen Convergenz" is employed by Schoenflies, see *Bericht*, p. 226, who uses the definition given by Osgood, *American Journal of Math.*, vol. XIX, p. 166. The term "Convergence Function" is also that employed by Schoenflies.

limits have for their limiting values two* numbers β_P^+, β_P^-, which may be called the *measures of non-uniform convergence at P on the right, and on the left*, respectively. If $\beta_P^+ = 0$, $\beta_P^- > 0$, the point P may be said to be one of uniform convergence on the right; a corresponding definition holds for the left. The measure β_P is the greater of the two numbers β_P^+, β_P^-; and at a point of uniform convergence $\beta_P^+ = \beta_P^- = 0$.

THE DISTRIBUTION OF POINTS OF NON-UNIFORM CONVERGENCE.

350. Let $u_1(x) + u_2(x) + \ldots + u_n(x) + \ldots$ denote a series of continuous functions which converges everywhere in the interval (a, b) to the sum $s(x)$. The most general possible distribution of the points of non-uniform convergence of the series will be here investigated.

In the first place, it can be shewn that *the points, at which the measure of non-uniform convergence β exceeds any fixed positive number A, form a closed set.*

For, if P be a limiting point of this set, in any semi-circle with P as centre there are points on the x-axis at which $\beta > \sigma$, and therefore there are points within the semi-circle at which $|R(x, y)| > \sigma$; and since this is the case however small the radius of the semi-circle may be, it follows that P is itself a point at which $\beta > \sigma$.

Next, it will be shewn that *the closed set, for which $\beta > \sigma$, is non-dense in the interval (a, b).*

At any point $P, (x, y)$, let a straight line of length 2ρ be drawn parallel to the y-axis, with P as its middle point, and let $\omega(\rho)$ be the fluctuation of the function $R(x, y)$ in the line 2ρ. The function $\omega(\rho)$ is a continuous function of ρ, since $R(x, y)$ is continuous with respect to y. If P be in the boundary $y = 0$, it will be sufficient to take the straight line of length ρ within the rectangle. Let $\alpha_\sigma(x, y)$ be the upper limit of the values of ρ which are such that $\omega(\rho) \leqq \sigma$. The function $\alpha_\sigma(x, y)$ is defined for every point in the rectangle, and is essentially either positive or zero. Since $R(x, y) = s(x) - s(x, y)$, and since $s(x)$ is independent of y, it follows that the function $\alpha_\sigma(x, y)$ is the same as the corresponding function defined for $s(x, y)$ instead of for $R(x, y)$.

The function $s(x, y)$ being everywhere continuous with respect to y, and being also continuous with respect to x, for every value of y except zero, it follows from the theorem of § 243, that $\alpha_\sigma(x, y)$ is an upper semi-continuous function.

* These numbers are equivalent to Osgood's *indices* of a point (b^+, b^-) of which he gives a different definition. The definition in the text is given in the paper in the *Proc. Lond. Math. Soc.* already quoted.

Let P be a point of the boundary $y=0$, at which the minimum of $\alpha_\sigma(x, 0)$ is not zero; it can be shewn that the saltus of $|R(x, y)|$ at P is $\leqq 2\sigma$. To prove this, we observe that a neighbourhood pp' of P can be found such that α_σ is at every point of pp', greater than a fixed number η which is less than the minimum of α_σ at P. Let X, Y be any two points in the rectangle of base pp' and height η, and let Xm, Ym' be perpendicular to the x-axis. We have then

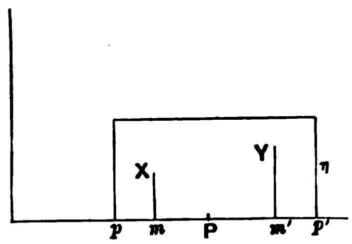


FIG. 5.

$$|R(X)-R(Y)| \leqq |R(X)-R(m)| + |R(Y)-R(m')|$$
$$\leqq 2\sigma,$$

and thus the required neighbourhood has been found.

It follows that, if the saltus of $|R(x, y)|$ at P be $>\sigma$, the minimum of $\alpha_{\frac{1}{2}\sigma}(P)$ with reference to the x-axis, must be zero. It has been shewn in § 184, that in every sub-interval of the x-axis, there are points at which this minimum of $\alpha_{\frac{1}{2}\sigma}(P)$ is positive. It follows that the closed set, for which $\beta > \sigma$, cannot be dense in any interval; and thus*:—

If a series of continuous functions converges to the sum $s(x)$ at every point of a given interval, then the points, at which the measure of non-uniform convergence exceeds a given positive number σ, form a non-dense closed set.

If we take a sequence of values of σ which converges to zero, we see that the set of all the points of non-uniform convergence of the series is the limit of the sequence of the closed non-dense sets which correspond to the values of σ. It follows that:—

The points of non-uniform convergence of a series of continuous functions which converge in a given interval to the sum $s(x)$, form in general a set of points of the first category; and the points of uniform convergence form a set of the second category, which is consequently everywhere-dense, and of the power of the continuum.

It is clear that the set of points at which the convergence function is indefinitely great, when it exists, forms a closed non-dense set.

If the functions $u_1(x)$, $u_2(x)$, $u_3(x)$, ... of which the sum in the interval (a, b) is the function $s(x)$, be discontinuous functions, there may be points of non-uniform convergence dependent upon the discontinuities of the given functions; and then it is no longer necessarily true that the points of uniform convergence are everywhere-dense.

* This theorem was given by Osgood for the case in which $s(x)$ is continuous, see *American Journal of Math.*, vol. XIX, 1897. The proof in the text was given by Hobson, *Proc. Lond. Math. Soc.*, vol. XXXIV, p. 245, also *Acta Mathematica*, vol. XXVII, p. 212.

EXAMPLE.

Let* $u_1(x)=0$, at all points of the interval (0, 1) except at the point $x=\frac{1}{2}$, where $u_1(x)=1$. Let $u_2(x)=-1$, at $x=\frac{1}{2}$, and $u_2(x)=1$, at $x=\frac{1}{4}, \frac{3}{4}$, and $u_2(x)=0$, everywhere else; let $u_3(x)=-1$ at $x=\frac{1}{4}, \frac{3}{4}$, and $u_3(x)=1$, at $x=\frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}$, and $u_3(x)=0$, at all other points; and so on. Then $s_1(x)$ is zero except at $x=\frac{1}{2}$, where $s_1(\frac{1}{2})=1$; $s_2(x)$ is zero except that $s_2(\frac{1}{4})=s_2(\frac{3}{4})=1$; $s_3(x)$ is zero except that $s_3(\frac{1}{8})=s_3(\frac{3}{8})=s_3(\frac{5}{8})=s_3(\frac{7}{8})=1$, and so on. The function $s(x)$ is everywhere zero, and therefore continuous in (0, 1); but the series everywhere converges and is non-uniformly convergent at every point of the interval (0, 1), since, in the neighbourhood of every assigned point, there are discontinuities of measure equal to 1, of $R_n(x)$.

THE LIMITS OF A SUM-FUNCTION AT A POINT.

351. Let α be a limiting point of the domain of the variable x, for which the convergent series $\Sigma u(x)$ is defined; α may, or may not, itself be a point of the domain. Let us suppose further that the limits $u_1(\alpha+0)$, $u_2(\alpha+0)$, ... $u_n(\alpha+0)$, ... at α on the right, all have definite values; it follows that $s_1(\alpha+0)$, $s_2(\alpha+0)$, ... $s_n(\alpha+0)$, ... also exist and have definite values.

We propose to examine the circumstances under which the limit $s(\alpha+0)$, of $s(x)$ on the right at α, exists, and the series $\Sigma u(\alpha+0)$ is convergent, with $s(\alpha+0)$ for its sum.

Let us assume that an interval $(\alpha, \alpha+\delta)$ exists, such that, for that part of the domain of x which falls within it, the given series is at least simply-uniformly convergent; and let us consider the transformed sum-function $s(x, y)$. If ϵ be an arbitrarily chosen positive number, a value y' of y can be found such that, for every value of x interior to $(\alpha, \alpha+\delta)$ which belongs to the domain, $|s(x, 0) - s(x, y')| < \epsilon$. With this value of y', an interval $(\alpha, \alpha+\eta)$, where $\eta \leqq \delta$, can be found, such that the fluctuation of $s(x, y')$ in the interval is $< \epsilon$, since the limit $s(\alpha+0, y')$ exists. If P, Q be any two points within the interval $(\alpha, \alpha+\eta)$ on the x-axis, and P', Q' are the points on the line $y=y'$, with the same values of x, we have

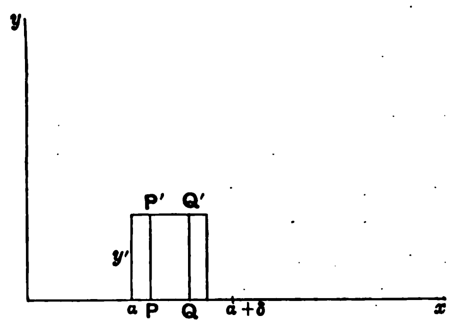


FIG. 6.

$$s(P) - s(Q) = [s(P') - s(Q')] + [s(P) - s(P')] + [s(Q') - s(Q)],$$

it follows that $|s(P) - s(Q)| < 3\epsilon$; and since ϵ is arbitrarily chosen, we see that $s(\alpha+0)$ has a definite value. The following theorem has therefore been established:—

* See W. H. Young, *Proc. Lond. Math. Soc.*, ser. 2, vol. II, p. 94.

If α be a limiting point of the domain of x for which the convergent series $\Sigma u(x)$ of which the sum-function is $s(x)$ is defined, and if, in a certain neighbourhood of α, on the right, the series converge at least simply-uniformly, then the function $s(x)$ has a definite limit $s(\alpha+0)$ at α, on the right. It is here assumed that $s_n(\alpha+0)$ has a definite value for each value of n.

This theorem specifies a sufficient condition for the existence of $s(\alpha+0)$; but the fulfilment of the condition does not ensure that the series $\sum_{n=1}^{\infty} u_n(\alpha+0)$ is convergent.

352. Necessary and sufficient conditions will now be determined, that $s(\alpha+0)$ may exist, and that the series $\Sigma u(\alpha+0)$ may converge to the value $s(\alpha+0)$. If $s(x, y)$ denotes the transformed sum-function, then the required conditions are those that the two limits $\lim_{x=\alpha}\lim_{y=0} s(x, y)$, $\lim_{y=0}\lim_{x=\alpha} s(x, y)$ should both exist, and should have the same value. The required conditions may consequently be obtained by applying the theorem of § 234, relating to repeated limits. We thus obtain the following theorem:—

The necessary and sufficient condition that the sum $s(x)$ of the convergent series $\Sigma u(x)$ may have a definite limit $s(\alpha+0)$ at the limiting point α of the domain of x, and that also the series $\Sigma u(\alpha+0)$, of which the terms are assumed to have definite values, may converge to the limit $s(\alpha+0)$, is that, corresponding to each arbitrarily chosen positive number ϵ, and to each integer n which is greater than some fixed number n_1 dependent on ϵ, a number θ can be found, such that, for every value of x belonging to the given domain and interior to the interval $(\alpha, \alpha+\theta)$, the condition $|R_n(x)| < \epsilon$ is satisfied, the number θ being in general dependent upon n.

In the particular case in which θ is, for each value of ϵ, independent of n, the point α is a point of uniform convergence on the right, of the series $\Sigma u(x)$, the point α itself being supposed to be excluded from the domain of x for which the series is defined; therefore *uniform convergence at α on the right is a sufficient condition that $s(\alpha+0)$ may have a definite value, and that the series $\Sigma u(\alpha+0)$ may converge to $s(\alpha+0)$; the point α being itself excluded, for the purpose, from the domain of x.*

By employing the alternative set of conditions for the existence and equality of repeated limits, given in § 235, we obtain the following theorem:—

The necessary and sufficient conditions that the sum $s(x)$ of the convergent series $\Sigma u(x)$ may have a definite limit $s(\alpha+0)$ to which the series $\Sigma u(\alpha+0)$ may converge, the terms of this series being assumed to have definite values, are (1) that $s_n(\alpha+0)$ should converge to a definite limit as n is indefinitely increased, and (2) that, corresponding to each arbitrarily chosen positive *number* ϵ, *and to each integer* n_1 *there should exist a value of* $n > n_1$, *and also a number* θ, *such that* $|R_n(x)| < \epsilon$ *for every value of* x *within the interval* $(a, a+\theta)$.

This theorem contains the completion and generalization of that of § 351. It is clear that the condition stated in the latter theorem is insufficient, without postulating that the condition (1) of the above theorem is satisfied, to ensure the existence and equality of the two repeated limits. When the conditions stated in the theorem are not satisfied, either or both of the limits $s(a+0)$, $\Sigma u(a+0)$ may exist; but they cannot both exist, and at the same time have one and the same value.

THE NECESSARY AND SUFFICIENT CONDITIONS FOR THE CONTINUITY OF THE SUM-FUNCTION.

353. If the functions $u_1(x)$, $u_2(x)$, ... $u_n(x)$, ... be all continuous throughout the domain of x, which will be taken to be the continuous interval (a, b), it has been shewn that a sufficient condition for the continuity of the sum-function $s(x)$ at a point x_1 is that a neighbourhood of x_1 can be found within which the series converges simply-uniformly; it has however been shewn, by means of examples, that this condition is not necessary for continuity of $s(x)$ at x_1.

The theorem of § 352 may be applied to obtain the necessary and sufficient condition for the continuity of $s(x)$ at x_1. That theorem shews that, in order that $s(x)$ may be continuous at x_1 on the right, it is necessary and sufficient that, corresponding to any arbitrarily chosen ϵ, an integer n_1 should exist, such that, for each n which is $> n_1$, a neighbourhood $(x_1, x_1+\theta)$ can be found, θ depending on n, such that $|R_n(x)| < \epsilon$, for every point x within this neighbourhood; where the integer n_1 may be so chosen that $|R_n(x_1)| < \epsilon$. A corresponding condition is necessary and sufficient to ensure that $s(x)$ is continuous at x_1 on the left. We can therefore state the necessary and sufficient condition of continuity at x_1 as follows:—

In order that $s(x)$ *may be continuous at the point* x_1, *it is necessary and sufficient that, corresponding to each arbitrarily chosen positive* ϵ, *a number* n_1 *can be found such that for each value of* $n > n_1$, *a neighbourhood* $(x_1-\delta_1, x_1+\delta_2)$ *of* x_1 *can be found, at every point of which* $|R_n(x)| < \epsilon$, *the numbers* δ_1, δ_2 *being dependent in general upon* n.

For a prescribed ϵ there is a certain range of values of y from zero upwards, for which $|R(x, y)| < \epsilon$; and the upper limit of these values of y may be denoted by $\phi_\epsilon(x)$: but there may be other greater values of y not continuous with the interval $(0, \phi_\epsilon(x))$, for which the condition $|R(x, y)| < \epsilon$, is also satisfied. At a point x_1 of non-uniform convergence of the series, the lower limit of $\phi_\epsilon(x)$, for the values of x in any neighbourhood of x_1, is zero, provided ϵ be chosen sufficiently small; whereas, for a point x_1 of uniform convergence, a neighbourhood of x_1 can be found for which the lower limit of $\phi_\epsilon(x)$ is greater than zero.

The second theorem of § 352, shews that, in the statement of the theorem, when ϵ and n_1 have been arbitrarily chosen, it is necessary and sufficient that a single integer $n > n_1$ should exist, and also a neighbourhood $(x_1 - \delta_1, x_1 + \delta_2)$, at every point of which $|R_n(x)| < \epsilon$, the numbers δ_1, δ_2 being dependent upon n. This is an alternative form of the theorem just stated.

The distinction between the three classes of points in the interval (a, b), viz. (1) those at which the series is uniformly convergent, (2) those at which the series is non-uniformly convergent, but at which the sum-function is continuous, and (3) those points at which the function is discontinuous, may be illustrated by means of figures* which indicate the regions of (x, y) in the neighbourhood of $(x_1, 0)$, at which $|R(x, y)|$ is less than an arbitrarily chosen ϵ.

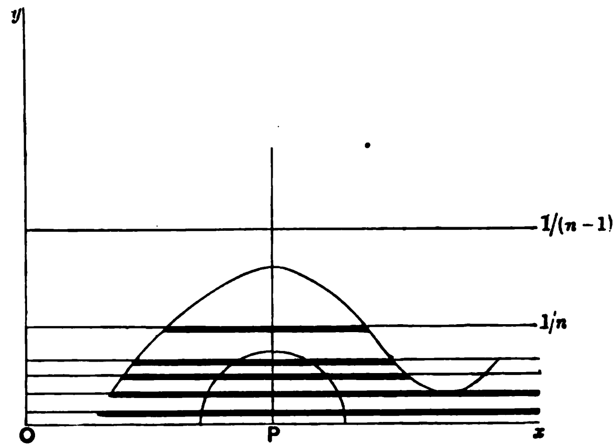


FIG. 7.

Fig. 7 represents the neighbourhood of a point P at which the convergence of the series is uniform. The blackened lines represent those portions of the lines whose ordinates are $1/n$, $1/(n+1)$, $1/(n+2)$, ... at which $|R_n(x)|$, $|R_{n+1}(x)|$ ... are $\leqq \epsilon$. These portions consist of all those parts of the lines which are bounded by the curve $y = \phi_\epsilon(x)$, there being also possibly such pieces outside the curve. An area, for example semi-circular, can be drawn, bounded by a portion of the x-axis containing P, and such that for every point within it $|R(x, y)| < \epsilon$; and that this should be possible for every value of ϵ is the condition that $R(x, y)$ be continuous at the point P with regard to the two-dimensional continuum (x, y).

* See Hobson, "On modes of convergence of an infinite series of functions of a real variable," *Proc. Lond. Math. Soc.*, ser. 2, vol. I.

Fig. 8 represents the neighbourhood of a point P at which the function $s(x)$ is continuous, but at which the series is non-uniformly convergent. In this case the function $\phi_\epsilon(x)$ is for all values of $\epsilon < \epsilon_0$, discontinuous at P. The

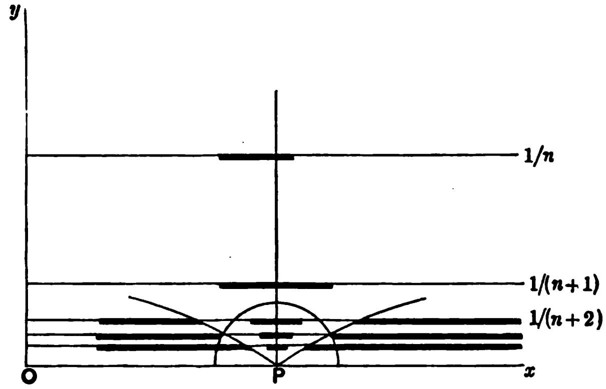


FIG. 8.

value of $\phi_\epsilon(x)$ at P is itself finite; but the functional limits $\phi_\epsilon(x_1+0)$, $\phi_\epsilon(x_1-0)$ at P are both zero. The breadth of the blackened portions of the straight lines parallel to the x-axis, which represent the portions of those lines at which $|R_n(x)| \leqq \epsilon$, diminishes indefinitely as y approaches the value zero at P. In this case no semi-circle can be drawn with P as centre, for all internal points of which $|R(x, y)| < \epsilon$; and thus the point P is one of non-uniform continuity, the measure of non-uniform convergence being ϵ_0. In the figure, the convergence is non-uniform on both sides of P; it is clear however in what manner the figure must be modified for the case in which the convergence is non-uniform on one side only of P. In case the measure of non-uniform convergence be indefinitely great the figure will be essentially similar to the above figure, whatever value of ϵ be chosen; otherwise the figure applies to an ϵ which is less than the measure ϵ_0 of non-uniform convergence, viz. the saltus at P of $|R(x, y)|$ in the two-dimensional continuum.

Fig. 9 represents the neighbourhood of a point P at which $s(x)$ is

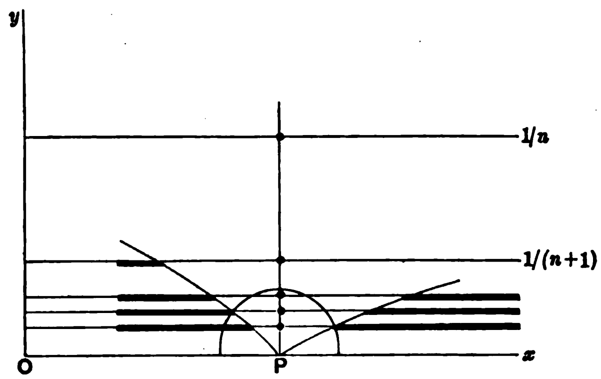


FIG. 9.

discontinuous, the value of ϵ being less than the measure of non-uniform convergence of the series at P. In this case, as before, $\phi_\epsilon(x)$ is finite at P, and $\phi_\epsilon(x_1+0)$, $\phi_\epsilon(x-0)$ are zero; but, on the parallels to Ox intersecting the ordinate at P, there are no intervals near P intersecting the ordinate, at which $|R(x, y)| < \epsilon$, but only points on the ordinate through P itself.

EXAMPLE.

As an example we may take the case in § 348, Ex. 3, $R_n(x)=\frac{nx}{1+n^2x^2}$, and thus

$$R(x, y)=\frac{xy}{x^2+y^2};$$

and we may suppose the domain of x to be the interval (0, 1). In this case, the point $x=0$ is a point of discontinuity of $R(x, y)$, and we find that if $\epsilon<\frac{1}{2}$, the condition

$$|R(x, y)|<\epsilon,$$

is satisfied for the space bounded by the x-axis, and by the straight line

$$y=\phi_\epsilon(x)=x\left[\frac{1}{2\epsilon}-\left(\frac{1}{4\epsilon^2}-1\right)^{\frac{1}{2}}\right].$$

The same condition is also satisfied for the space between the y-axis and the straight line

$$y=x\left[\frac{1}{2\epsilon}+\left(\frac{1}{4\epsilon^2}-1\right)^{\frac{1}{2}}\right],$$

and thus the point $x=0$ is a point of continuity of the function $s(x)$, although the convergence is non-uniform at that point. If $\epsilon>\frac{1}{2}$, then $|R(x, y)|<\epsilon$, for the whole space between the axes; and thus the measure of non-uniform convergence at the point $x=0$ is $\frac{1}{2}$.

354. The necessary and sufficient conditions will now be determined that $s(x)$ may be continuous in the whole interval (a, b). First let us assume that $s(x)$ is everywhere continuous. Choose an arbitrarily small positive number ϵ; then, if n be sufficiently large, there are points in (a, b) at which $|R_n(x)| < \epsilon$. If P be such a point, a neighbourhood of P can be found within which the condition $|R_n(x)| < \epsilon$ is everywhere satisfied; the size of this neighbourhood may be extended in both directions until points p, q are reached, at which $|R_n(p)| = |R_n(q)| = \epsilon$; for this follows from the fact that $R_n(x)$ is a continuous function of x. Every such point P in (a, b), at which $|R_n(x)| < \epsilon$, may, in a similar manner, be enclosed in an interval of finite length; and in all internal points of such intervals the condition $|R_n(x)| < \epsilon$, is satisfied. Thus, for any fixed value of n which is sufficiently large, there exists a finite, or infinite, set D_n of intervals in (a, b) which do not overlap, such that, at every point which is interior to one of the intervals of the set D_n, the condition $|R_n(x)| < \epsilon$, is satisfied; moreover, the intervals contain in their interiors all points except a, b, at which the condition is satisfied. Two intervals of D_n may abut on one another at a point in which $|R_n(x)|$ is equal to ϵ; otherwise the intervals will be separated from one another.

Let us consider the systems of intervals $D_n, D_{n+1}, D_{n+2}, \ldots$; every value of n being taken from a fixed value onwards. The whole set thus formed is such that every point in (a, b), except the end-points, is interior to an infinite number of intervals of the compound set; this follows from the fact that, for any point x, a value of n, say n_1, can be found such that

$$|R_{n_1}(x)| < \epsilon, \ |R_{n_1+1}(x)| < \epsilon \ldots.$$

Moreover, intervals can be found with a, b as end-points, which for a sufficiently large value of m belong to D_{n+m}. In accordance with the Heine-Borel theorem, established in § 68, a finite set of intervals can be selected from the set which consists of $D_n, D_{n+1}, D_{n+2}, \ldots$, which contains every point of (a, b) as an internal point of one of the intervals at least, and such that a, b are end-points of two of the intervals of the finite set. It thus appears that, on the supposition that $s(x)$ is continuous in the whole interval (a, b), if n be any integer chosen arbitrarily, a finite set of numbers $n+\iota_1, n+\iota_2, \ldots n+\iota_r$ all greater than or equal to n, can be found, such that, for every value of x, $|R_m(x)| < \epsilon$, where m has one of the values $n+\iota_1, n+\iota_2, \ldots n+\iota_r$. The particular value of m varies with x, but the same value of m is applicable to the whole of one of a finite number of continuous intervals; also the set of values of m is dependent on the chosen ϵ. The intervals of the set will overlap; but an overlapping portion may be considered to belong to one of the intervals only, so that (a, b) may be divided into a finite number of parts, in each of which, for some value of m, constant for that part, the condition $|R_m(x)| < \epsilon$ is satisfied.

It should be remarked that the set D_n, for a fixed n, is not necessarily a finite set; for, besides those intervals for which $|R_n(x)| > \epsilon$, and the intervals D_n themselves, there may be points of (a, b) which are not in either set of intervals but are limiting points of end-points of the intervals D_n; and at such points $|R_n(x)| = \epsilon$. For example, if $|R_n(x)| = \frac{1}{n^2} + \frac{x-c}{n^3}\sin\left(\frac{1}{x-c}\right)$, where $a < c < b$, and if $\epsilon = 1/n^2$, the point c is a point of continuity of $R_n(x)$, and is a limiting point of end-points of those intervals for which $|R_n(x)| < \epsilon$.

Conversely, if for every value of ϵ a finite set of intervals exists, which has the property described above, the function $s(x)$ is continuous in (a, b). For let us consider a point P of non-uniform convergence of the series. Then for a given ϵ, P is inside an interval for every point of which, for a fixed value of m, $|R_m(x)| < \epsilon$, and it has been shewn that this is the condition that $s(x)$ may be continuous at P: hence every point of non-uniform convergence of the series is a point of continuity of $s(x)$. The condition has now been obtained in the following form:—

The necessary and sufficient condition for the continuity in (a, b) of the sum-function of a series $u_1(x) + u_2(x) + \ldots$, each term of which is a continuous *function of x throughout (a, b), and which converges at every point of this domain to a definite value $s(x)$, is that, corresponding to any arbitrarily chosen number ϵ, and to an arbitrarily chosen integer n, the condition $|R_m(x)| < \epsilon$ is satisfied for every value of x in (a, b), where m has one of a finite number of values all greater than or equal to n, the value of m depending in general on x, but being constant for all points x which lie in one of a number of finite portions of the interval (a, b).*

This theorem, which was first established by Arzelà*, states that a certain mode of convergence in the interval is the necessary and sufficient condition for the continuity of the sum-function; and this mode has been termed by Arzelà*, *convergenza uniforme a tratti* (uniform convergence by segments). The term is perhaps not altogether appropriate, because the intervals are dependent in number and length upon the arbitrarily chosen ϵ. Uniform convergence, and simple-uniform convergence, are special cases of this mode of convergence; for in these cases the finite set of intervals which corresponds to a given ϵ, reduces to one interval, viz. the whole interval (a, b).

EXAMPLES.

1. The series
$$1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\ldots+\frac{x^n}{n!}+\ldots$$

is convergent in any finite interval (a, b) whatever. It is shewn in elementary treatises that the series converges to e^x, for all rational values of x. In order to extend the proof of the exponential theorem to the case of an irrational value of x, we observe that the above series converges uniformly in the interval (a, b), since $\left|\frac{x^n}{n!}\right| < \frac{k^n}{n!}$, where k is a fixed number greater than $|a|$, and $|b|$; and hence, in accordance with the theorem of § 345, since $\Sigma\frac{k^n}{n!}$ is convergent, the given series converges uniformly in (a, b). It follows that the sum-function $s(x)$ of the series is continuous in (a, b). Further, the function e^x has been defined for an irrational value of x, by extension (see § 191) of the function as defined for rational values of x; and it was shewn in § 37, that the function e^x, so defined for the whole domain, is single-valued at the irrational points, and therefore it is continuous. The two functions e^x, $s(x)$ are both continuous in (a, b), and have identical values at the rational points; therefore, in accordance with the theorem of § 173, they are identical everywhere in (a, b). Therefore e^x is the sum-function of the series in any finite interval (a, b).

2. It is proved in elementary treatises that, for a value of x which is numerically less than unity, the binomial series

$$1+nx+\frac{n(n-1)}{2!}x^2+\ldots+\frac{n(n-1)\ldots(n-r+1)}{r!}x^r+\ldots$$

* See the memoir "Sulle serie di funzioni," *Mem. della R. Accad. degli Sci. di Bologna*, ser. 5, vol. VIII, 1900. The proof given above was published by Hobson in the *Proc. Lond. Math. Soc.*, ser. 2, vol. I, p. 380.

converges to a suitable value of $(1+x)^n$, when n is a rational number. To extend the theorem to the case in which n may have an irrational value, consider an interval (n_1, n_2) of n, where n_1 and n_2 are rational numbers.

We have $$\left|\frac{n(n-1)\dots(n-r+1)}{r!}x^r\right| < \frac{N(N+1)\dots(N+r-1)}{r!}|x|^r,$$

where N is the greater of the numbers $|n_1|$ and $|n_2|$. The number x remaining fixed, we thus see that, for all values of n in the interval (n_1, n_2), each term of the series is numerically less than the corresponding term of the convergent series

$$1+N|x|+\frac{N(N+1)}{2!}|x|^2+\dots;$$

therefore the series converges uniformly for all values of n in the interval (n_1, n_2). Hence the sum-function of the series, for a fixed value of x, is a continuous function of n in the interval (n_1, n_2). The function $(1+x)^n$ of n, was defined in § 37, for irrational values of n, by extension of the function considered as defined only for rational values of n; and it was shewn that the function so obtained by extension is single-valued, and it is therefore continuous. As in example (1), it now follows, that, for the fixed value of x, numerically <1, the sum of the series is for all values of n in (n_1, n_2) represented by the suitable value of $(1+x)^n$. The interval (n_1, n_2) is arbitrary.

THE CONVERGENCE OF POWER-SERIES.

355. A series of which the $(n+1)$th term is of the form $a_n x^n$ is called a power-series. It will be assumed that the domain of x is a continuous one.

If the power-series $a_0+a_1x+a_2x^2+\dots+a_nx^n+\dots$ be such that, for a positive value $x=X$, every term a_nx^n is numerically less than some fixed positive number A, the series converges absolutely for every value of x which is numerically less than X.

The partial remainder

$$R_{n,m}(x)=a_nx^n+a_{n+1}x^{n+1}+\dots+a_{n+m-1}x^{n+m-1}$$

is such that

$$\begin{aligned}|R_{n,m}(x)| &\leqq |a_nx^n|+|a_{n+1}x^{n+1}|+\dots+|a_{n+m-1}x^{n+m-1}|\\ &< \left(\frac{|x|}{X}\right)^n A\left\{1+\frac{|x|}{X}+\frac{|x|^2}{X^2}+\dots+\frac{|x|^{m-1}}{X^{m-1}}\right\}\\ &< A\frac{|x|^n}{X^n}\left(1-\frac{|x|}{X}\right)^{-1};\end{aligned}$$

hence, for any fixed value of x such that $|x|<X$, n may be so chosen that all the remainders $R_{n,m}(x)$ are numerically less than an arbitrarily chosen number; and therefore the series converges at the point x. It is also clear that the convergence is absolute.

If the series diverge for the value x of X, it diverges also for every value of x which is numerically greater than X.

For if the series converged for a value x_1 of x numerically greater than X, the condition of the preceding theorem would be satisfied by $|x_1|$, and hence the series would converge absolutely for the value X; which is contrary to the hypothesis.

The power-series may converge (1) for no value of x except zero, (2) for every value of x, or (3) for a value X of x different from zero, but not for every value of x.

In case (3), there exists a definite interval $(-R, R)$ such that the series converges for every value of x in the interior of the interval, and diverges for every value of x exterior to the interval. The series may, or may not, converge at either end-point of the interval.

The interval $(-R, R)$ is called *the interval of convergence* of the series. To establish the existence of this interval, we observe that, if the series converge for any value x_1 of x, it converges for every value numerically less than $|x_1|$, because, then, every term $a_n x_1^n$ is numerically less than some fixed number A. It has been shewn that, if the series diverge for any particular value of x, it diverges for all numerically greater values. Hence those numbers $|x|$ which are such that the series converges for $|x|$, must have an upper limit R, which must also be the lower limit of those values of $|x|$ for which the series diverges; and this limit R determines the interval $(-R, R)$ of convergence.

356. *If* the power-series converge for a value X of x, greater than zero, it converges uniformly in the interval $(-X_0, X)$, where $0 < X_0 < X$.*

We have

$$R_n(x) = R_{n,1}(X)\left(\frac{x}{X}\right)^n + \sum_{p=2}^{\infty}\{R_{n,p}(X) - R_{n,p-1}(X)\}\left(\frac{x}{X}\right)^{n+p-1}$$

$$= \sum_{p=1}^{\infty} R_{n,p}(X)\left(1-\frac{x}{X}\right)\left(\frac{x}{X}\right)^{n+p-1},$$

hence

$$|R_n(x)| < \epsilon\left(1-\frac{x}{X}\right)\sum_{1}^{\infty}\left|\frac{x}{X}\right|^{n+p-1} < \epsilon\,\frac{1-\frac{x}{X}}{1-\left|\frac{x}{X}\right|}\left|\frac{x}{X}\right|^{n}, \text{ if } |x| < X,$$

provided n be so chosen that $|R_{n,p}(X)| < \epsilon$, for every value of p; which is possible by reason of the convergence of the series for the value X of x. For such value of n, and for all greater values, $|R_n(x)| < \epsilon \cdot \frac{X+X_0}{X-X_0}$, for every point in the interval $(-X_0, X)$; hence the series converges uniformly in this interval, including the end-points.

It follows from this theorem that the sum-function $s(x)$ is continuous in the interval $(-X_0, X)$.

* See Abel's *Œuvres*, vol. I, p. 223.

In case the series be convergent at the point R at the extremity of the interval of convergence, we see from the theorem that the convergence is uniform in the interval $(-R_0, R)$, where $R_0 < R$; and that consequently the sum-function is continuous in this interval, and is continuous at the points R, and $-R_0$.

We have thus established the theorem due to Abel*, that:—

If the series $a_0 + a_1x + a_2x^2 + \ldots$, which converges within an interval of convergence of which R is one of the ends, be such that the series converges for $x = R$, then the sum of the series for $x = R$ is continuous with the sum-function for the interior of the interval of convergence.

This theorem may also be deduced† from the theorem in § 348. For we have $\left(\frac{x}{R}\right)^n \geqq \left(\frac{x}{R}\right)^{n+1}$, for $0 \leqq x \leqq R$, and for all values of n; hence, since the series $\sum_{n=0}^{\infty} a_n R^n$ is, by hypothesis, convergent, it follows that the series $\sum_{n=0}^{\infty} a_n x^n$ is uniformly convergent in the interval $(0, R)$. Therefore the sum-function is continuous in that interval, including the end-point R.

It should be observed that this theorem has been established only for a series in which the powers of the variable are ascending, and that it is not necessarily true in any other case. For example, the series $x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \ldots$ is convergent within the interval $(-1, 1)$; and as the series is, for such values of x, absolutely convergent, the series $x + \frac{1}{3}x^3 - \frac{1}{2}x^2 + \frac{1}{5}x^5 + \frac{1}{7}x^7 - \frac{1}{4}x^4 + \ldots$ has the same sum-function within the interval, that function being $\log_e(1+x)$. At $x = 1$, the series $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots$ is convergent, and in accordance with the theorem its sum is $\log_e 2$; but the series $1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \ldots$, although convergent, has the sum $\frac{3}{2}\log_e 2$ (see Ex. 1, § 334), which is not continuous with the sum of the series $x + \frac{1}{3}x^3 - \frac{1}{2}x^2 + \ldots$.

357. *If the series $a_0 + a_1x + a_2x^2 + \ldots$ converge within an interval $(-R, R)$, and be such that, in every interval $(-\delta, \delta)$, where δ is an arbitrarily chosen number $< R$, $s(x)$ vanishes for some value of x which is not zero, the coefficients $a_0, a_1, a_2, \ldots$ must all be zero.*

If δ have any value $< R$, the function $s(x)$ is continuous in the interval $(-\delta, \delta)$; hence, if ϵ be an arbitrarily chosen positive number, δ_1 may be chosen so small that $|s(x) - a_0| < \epsilon$, if x be in the interval $(-\delta_1, \delta_1)$; and by hypothesis there is in the interval one value of x such that $s(x) = 0$; therefore $|a_0| < \epsilon$, and since ϵ is arbitrary, we have $a_0 = 0$. The series $a_1 + a_2x + a_3x^2 + \ldots$ converges in the interval $(-\delta, \delta)$, and its sum vanishes for some value of x which is not zero, hence the same argument as before establishes that $a_1 = 0$; proceeding in this manner, it can be shewn that $a_2, a_3, \ldots$ vanish.

* Abel's *Œuvres*, *Crelle's Journal*, vol. I, p. 223; also Dirichlet, *Liouville's Journal*, ser. 2, vol. VII, p. 253.

† Hardy, *Proc. Lond. Math. Soc.*, ser. 2, vol. IV, p. 252.

It follows, as a corollary from this theorem, that *there cannot be two distinct power-series, each of which converges within some interval, and such that in every sub-interval* $(-\delta, \delta)$ *there is a point distinct from zero, at which the sum-functions have identical values.*

358. Let us suppose that all the series

$$a_{11} + a_{12}x + a_{13}x^2 + \ldots + a_{1r}x^{r-1} + \ldots,$$
$$a_{21} + a_{22}x + a_{23}x^2 + \ldots + a_{2r}x^{r-1} + \ldots,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$a_{n1} + a_{n2}x + a_{n3}x^2 + \ldots + a_{nr}x^{r-1} + \ldots,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

are, for every value of n, absolutely convergent at the point $x = R$; each series is then both absolutely and uniformly convergent in the interval $(-R, R)$. Let us denote by $u_n(x)$ the sum of the series $a_{n1} + a_{n2}x + \ldots + a_{nr}x^{r-1} + \ldots$ at a point x in the interval, and let U_n denote the sum of the series

$$|a_{n1}| + |a_{n2}R| + \ldots + |a_{nr}R^{r-1}| + \ldots.$$

If the series $$U_1 + U_2 + \ldots + U_n + \ldots$$

be convergent, the series

$$u_1(x) + u_2(x) + \ldots + u_n(x) + \ldots$$

is, in accordance with § 345, uniformly convergent in the interval $(-R, R)$, the end-points included; it follows that $s(x)$, the sum of this series, is continuous in the interval $(-R, R)$.

The series $u_1(x) + u_2(x) + \ldots + u_n(x) + \ldots$ can be arranged as a series of type ω^2, by substituting for $u_1(x), u_2(x), \ldots$ the various series in powers of x. Moreover this series is absolutely convergent; for the terms of the series

$$|a_{11}| + |a_{12}x| + \ldots + |a_{1r}x^{r-1}| + \ldots + |a_{21}| + |a_{22}x| + \ldots$$

are each less than the corresponding terms of the series obtained by writing R for x; and the latter series is $U_1 + U_2 + \ldots$, which is convergent.

Since the series $u_1(x) + u_2(x) + \ldots$ is absolutely convergent when the power-series are substituted for $u_1(x), u_2(x), \ldots$, it remains (see § 335) absolutely convergent when it is arranged in the form

$$b_1 + b_2x + b_3x^2 + \ldots + b_rx^{r-1} + \ldots,$$

where $$b_1 = a_{11} + a_{21} + a_{31} + \ldots,$$
$$b_2 = a_{12} + a_{22} + a_{32} + \ldots,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$b_r = a_{1r} + a_{2r} + a_{3r} + \ldots,$$

and its sum is unaltered. It has thus been shewn that the continuous function $s(x)$ can be represented in the closed interval $(-R, R)$ by the power-series

$$b_1 + b_2 x + b_3 x^2 + \ldots + b_r x^{r-1} + \ldots.$$

The following theorem has therefore been established:—

If $u_1(x)$, $u_2(x)$, ... be functions which can be represented by power-series that are all absolutely convergent at the point R, and therefore in the interval $(-R, R)$, and if the series $u_1(x) + u_2(x) + \ldots$ be absolutely convergent at $x = R$, then the series $u_1(x) + u_2(x) + \ldots$ converges in the interval to a continuous sum-function $s(x)$, which is the sum of the power-series obtained by substituting the different power-series for the terms $u_1(x)$, $u_2(x)$, ..., and rearranging the resulting series.

359. Let $s(x)$ denote the sum of the power-series $a_1 + a_2 x + a_3 x^2 + \ldots$ which converges at all points interior to the interval $(-R, R)$; then, provided x, $x + h$ be both interior to the interval, the series

$$a_1 + a_2(x+h) + a_3(x+h)^2 + \ldots + a_n(x+h)^{n-1} + \ldots$$

converges absolutely to the sum $s(x+h)$; and if we expand each term $(x+h)^{n-1}$ in powers of h, we have for each n a finite series of powers of h which may be regarded as absolutely convergent. In accordance with the theorem of § 358, if we arrange the series in powers of h,

$$(a_1 + a_2 x + a_3 x^2 + \ldots) + (a_2 + 2a_3 x + \ldots + \overline{n-1}\, a_n x^{n-2} + \ldots)\, h + \ldots,$$

this series will represent a continuous function of h, provided $x, x+h$ lie in an interval $(-r, r)$, where $r < R$; and the sum-function is $s(x+h)$. We have then

$$\frac{s(x+h) - s(x)}{h} = \{a_2 + 2a_3 x + \ldots + (n-1)\, a_n x^{n-2} + \ldots\} + h v_1(x) + h^2 v_2(x) + \ldots,$$

where $v_1(x)$, $v_2(x)$ are continuous functions of x. As h converges to zero, the convergent series $h v_1(x) + h^2 v_2(x) + \ldots$, of which the sum is a continuous function of h, converges to zero; therefore $\frac{s(x+h) - s(x)}{h}$ has as its limit, when h is indefinitely diminished, the sum of the convergent series

$$a_2 + 2a_3 x + \ldots + (n-1)\, a_n x^{n-2} + \ldots;$$

and thus the function $s(x)$ possesses a differential coefficient, which is the sum of the convergent series obtained by differentiating the terms of the power-series which represents $s(x)$. We have thus obtained the following theorem:—

If $s(x)$ be the sum of a power-series which converges within a given interval, the function $s(x)$ has a differential coefficient $s'(x)$ at each point x within the interval of convergence; and the series obtained by means of a term by term differentiation of the given series converges at such a point to the sum $s'(x)$.

[illegible]

[illegible]

[illegible]

[illegible]

[illegible] convergence [illegible] following [illegible]—

[illegible] convergent [illegible]

[illegible]

We have

[illegible]

[illegible]

[illegible] number [illegible] We have then

[illegible]

[illegible] there exists a fixed number [illegible]

[illegible]

* [illegible]

The numbers m, n can be chosen so great that $|s_n s_n' - ss'| < \theta$, where θ is arbitrarily fixed, and also such that $|\Sigma_n' - \Sigma'_{n-m+1}|$ is less than θ', where θ' is arbitrarily chosen; and if this be done we have

$$|S_n - ss'| < \theta + \eta(\Sigma' - \Sigma_1') + A\theta'.$$

But θ, η, θ' are each arbitrarily small; hence n can be so chosen that $|S_n - ss'|$ is arbitrarily small; and thus the series Σc converges to the sum ss'.

361. In case both the series Σa, Σb be only conditionally convergent, we are unable to assert that the series Σc converges; but in case it do converge, its sum is given by the following theorem due to Abel*:—

In case the product series Σc of two convergent series Σa, Σb be itself also convergent, its sum is the product ss' of the sums of the two given series.

Since the series

$$a_1 + a_2 + \ldots + a_n + \ldots, \quad b_1 + b_2 + \ldots + b_n + \ldots$$

are both convergent, the series

$$a_1 + a_2 x + a_3 x^2 + \ldots + a_n x^{n-1} + \ldots,$$

$$b_1 + b_2 x + b_3 x^2 + \ldots + b_n x^{n-1} + \ldots$$

converge absolutely for all values of x, such that $0 \leqq x < 1$; this follows from the theorem established in § 355. The product series formed from these two series is

$$c_1 + c_2 x + c_3 x^2 + \ldots + c_n x^{n-1} + \ldots;$$

and this, in accordance with Cauchy's theorem as to the product of two absolutely convergent series, is convergent for $0 \leqq x < 1$, and converges to $s(x)s'(x)$. Since the series is by hypothesis convergent when $x = 1$, its sum for this value is (see § 356) continuous with its sum, $s(x)s'(x)$, for all values of x which are numerically < 1; also $\lim_{x=1} s(x) = s$, $\lim_{x=1} s'(x) = s'$, and hence the series Σc converges to the value ss'.

It will be observed that the above theorem of Abel does not give any criterion which decides whether the product of two conditionally convergent series is, or is not, convergent. Criteria applicable to special classes of cases have been given by Pringsheim†.

TAYLOR'S SERIES.

362. If a function $f(x)$ be such that, at every point within the interval $(-R, R)$, it is the sum of the convergent series

$$a_1 + a_2 x + a_3 x^2 + \ldots + a_n x^{n-1} + \ldots,$$

* *Crelle's Journal*, vol. I, also *Œuvres*, vol. I, p. 226.

† *Math. Annalen*, vol. XXI, "Ueber die Multiplication bedingt convergenter Reihen."

it has been shewn in § 359 that $f'(x)$ exists at every point within the interval, and that it is the sum of the convergent series

$$a_2 + 2a_3x + \ldots + (n-1)\, a_n x^{n-2} + \ldots.$$

A second application of the same theorem shews that $f''(x)$ exists, and is the sum of the convergent series

$$1 \, . \, 2a_3 + 2 \, . \, 3a_4 x + \ldots + (n-1)(n-2)\, a_n x^{n-3} + \ldots.$$

Proceeding in this manner, it can be shewn that $f^{(r)}(x)$ exists, for every value of r, at every point x interior to the interval $(-R, R)$, and that the series

$$1 \, . \, 2 \, . \, 3 \ldots r a_{r+1} + 2 \, . \, 3 \ldots (r+1)\, a_{r+2} x + \ldots$$

converges to the value $f^{(r)}(x)$.

It has further been shewn that, if $x+h$ also lies within the interval $(-R, R)$, the series obtained by arranging the series

$$a_1 + a_2(x+h) + a_3(x+h)^2 + \ldots$$

as a series in powers of h converges to the value $f(x+h)$. The coefficient of h^r in this series is

$$a_{r+1} + (r+1)\, a_{r+2} x + \frac{(r+2)(r+1)}{2!} a_{r+3} x^2 + \ldots,$$

which is $\frac{1}{r!} f^{(r)}(x)$. It has thus been shewn that the series

$$f(x) + hf'(x) + \frac{h^2}{2!} f''(x) + \ldots + \frac{h^r}{r!} f^{(r)}(x) + \ldots$$

converges to the value $f(x+h)$, provided x, $x+h$ be both interior to the interval $(-R, R)$ of convergence of that power-series of which $f(x)$ is the sum.

This theorem is a particular case of Taylor's theorem for the expansion of a function $f(x+h)$ in powers of h, and has here been established for the particular case of a function $f(x)$ which represents the sum of a convergent power-series. It has moreover been proved that such a function possesses differential coefficients of all orders within the interval of convergence of the power-series.

We proceed to investigate the necessary and sufficient conditions that a corresponding theorem may hold for functions which are not defined by means of a power-series.

363. Let $f(x)$ be a function defined for the interval $(a, a+\lambda)$, where λ is a positive number, and satisfying the conditions (1) that, at the point a, the first $n-1$ derivatives of $f(x)$ on the right all exist; (2) that at every point x such that $a < x < a+\lambda$, the first $n-1$ differential coefficients of $f(x)$ all exist, and are continuous, being also continuous with the derivatives on the right at a; (3) that $f^{(n)}(x)$ exists at each point in the interior of the interval $(a, a+\lambda)$, having values which are finite, or infinite, with fixed sign.

Let the number K be defined by the equation

$$f(a+h)-f(a)-hf'(a)-\frac{h^2}{2!}f''(a)-\dots-\frac{h^{n-1}}{(n-1)!}f^{(n-1)}(a)=h^{n-\nu}K,$$

where h is such that $0<h<\lambda$, and ν is a fixed positive or negative integer, or zero, but such that $n-\nu$ is positive; the derivatives $f'(a), f''(a), \dots$ are those of $f(x)$ on the right at a.

Next, let $F(x)$ denote the function

$$f(a+h)-f(x)-(a+h-x)f'(x)-\frac{(a+h-x)^2}{2!}f''(x)-\dots$$
$$-\frac{(a+h-x)^{n-1}}{(n-1)!}f^{(n-1)}(x)-(a+h-x)^{n-\nu}K,$$

where K has the value defined above, and x is in the interval $(a, a+h)$. The function $F(x)$ is continuous in the interval, and $F'(x)$ exists everywhere in the interior of the interval. Moreover, since $F(x)$ vanishes for $x=a$, and for $x=a+h$, it follows from the theorem of § 203 that $F'(x)$ vanishes for some value of x within the interval $(a, a+h)$; let this value be $a+\theta h$, where θ is a number such that $0<\theta<1$.

Since

$$F'(x)=-\frac{(a+h-x)^{n-1}}{(n-1)!}f^{(n)}(x)+(n-\nu)(a+h-x)^{n-\nu-1}K,$$

we see that
$$K=\frac{h^\nu(1-\theta)^\nu}{(n-\nu)(n-1)!}f^{(n)}(a+\theta h);$$

therefore, from the definition of K, we have

$$f(a+h)=f(a)+hf'(a)+\frac{h^2}{2!}f''(a)+\dots+\frac{h^{n-1}}{(n-1)!}f^{(n-1)}(a)$$
$$+\frac{h^n(1-\theta)^\nu}{(n-\nu)(n-1)!}f^{(n)}(a+\theta h).$$

It is clear that a corresponding result holds for an interval on the left of the point a, provided corresponding conditions hold as to the existence of the differential coefficients; the derivatives at a being in this case those on the left.

In case $f(x)$ be defined for an interval $(a-\lambda', a+\lambda)$, and the first $n-1$ differential coefficients of $f(x)$ exist at every point x in the interior of the interval, and $f^{(n)}(x)$ everywhere exist in the interval of which the end-points are $a, a+h$, where h is any number such that $-\lambda'<h<\lambda$, the theorem holds for every such value of h, positive or negative; $f'(a), f''(a), \dots$ denoting in this case the differential coefficients at a.

This theorem is frequently spoken of as Taylor's theorem, although that name was originally, and is still usually, applied to the case in which it is possible to suppose n to be indefinitely increased, so that the series becomes an infinite convergent one.

The expression $R_n = \dfrac{h^n (1-\theta)^\nu}{(n-\nu)(n-1)!} f^{(n)}(a+\theta h)$, where ν is a positive or negative integer such that $n-\nu > 0$, is spoken of as "the remainder in Taylor's series." In this general form it was obtained by Schlömilch* and by Roche†. The particular case in which $\nu = 0$, $R_n = \dfrac{h^n}{n!} f^{(n)}(a+\theta h)$, is known as Lagrange's form‡ of the remainder in Taylor's series; another particular case, due originally to Cauchy§, of the general form given by Schlömilch, is that in which $\nu = n-1$, or $R_n = \dfrac{h^n (1-\theta)^{n-1}}{(n-1)!} f^{(n)}(a+\theta h)$.

364. If $f(x)$ possess differential coefficients of all orders within a prescribed interval $(a-\lambda', a+\lambda)$, then, provided R_n have the limit zero, when n is indefinitely increased, for each value of h, the series

$$f(a) + hf'(a) + \frac{h^2}{2!} f''(a) + \ldots + \frac{h^n}{n!} f^{(n)}(a) + \ldots,$$

where $-\lambda' < h < \lambda$, is convergent, and has $f(a+h)$ for its limiting sum. This is Taylor's theorem in the original sense of the term.

It will be observed that the existence of differential coefficients at the extreme points $-\lambda'$, λ has not been presupposed, but only their existence for all points for which $-\lambda' < h < \lambda$. If the condition $\lim\limits_{n=\infty} R_n = 0$ be satisfied for each value of h within the interval $(-\lambda', \lambda)$, and if the series converge also for $h = \lambda$, then, since it is a power-series, it follows from the theorem of § 356, that at $h = \lambda$ the series converges to the value $f(a+\lambda)$.

The value of θ, in any of the forms of the expression for R_n, is in general dependent upon n; and consequently it is not a sufficient condition of convergence of the series that R_n have the limit zero as n is indefinitely increased, whilst θ remains fixed, even though this be the case for each fixed value of θ in the interval (0, 1). In connection with the theory of non-uniform convergence of series we have already seen in § 349, that a function $R_n(x)$ may have the limit zero, as n is increased indefinitely, for each fixed value of x in a given interval, and yet that $\lim R_n(x)$ may not necessarily be zero when x varies with n.

* *Handbuch der Differential- und Integralrechnung*, 1847.

† *Mém. de l'Acad. de Montpellier*, 1858. See also *Liouville's Journal*, ser. 2, vol. III, pp. 271 and 384.

‡ *Théorie des Fonctions*, vol. I, p. 40.

§ *Calcul Diff.*, p. 77.

For example, if $R_n = \frac{n\theta h}{(1+\theta h)^n}$, then R_n has the limit zero for every fixed value of θ; but if $\theta = 1/n$, R_n has the limit he^{-h}.

A sufficient condition for the convergence of the series is that R_n, for each fixed value of h within the given interval, as n is indefinitely increased, should converge to zero uniformly for all values of θ in the interval (0, 1). Thus, for each value of h, and each value of an arbitrarily chosen positive number ϵ, a value n_1 of n, would exist such that

$$\left|\frac{h^n(1-\theta)^\nu}{(n-\nu)(n-1)!} f^{(n)}(a+\theta h)\right| < \epsilon,$$

provided $n \geqq n_1$, for every value of θ in the interval (0, 1).

This condition, though sufficient for the convergence of the series, has not been shewn to be necessary. An investigation, due to Pringsheim*, will now be given of the necessary and sufficient conditions for the convergence of Taylor's series.

365. *If the series* $\sum_0^\infty c_n h^n$ *converge for every positive value of* h *which is* $< R$, *and if* $f(x)$ *denote the sum of the series* $\sum_0^\infty c_n (x-a)^n$, *where* a *is a fixed number, and* $0 \leqq x - a < R$, *then* (1) $f(x)$ *possesses for every value of* x, *such that* $a \leqq x < a + R$, *a definite finite value; and* (2) *for every* x, *such that* $a < x < a + R$, $f(x)$ *possesses finite differential coefficients of every order, and at* a, *derivatives on the right of every order; also* (3) *the condition is satisfied that* $\frac{1}{(n\pm p)!} f^{(n)}(a+h) \,.\, k^{n\pm p}$ *converges uniformly for all values of* h, k *such that* $0 \leqq h \leqq h + k \leqq r$, *to zero, as* n *is indefinitely increased, where* $r < R$, *and* p *is any arbitrarily chosen integer, which may be zero.*

This theorem contains necessary conditions for the convergence of Taylor's series for an interval on the right of a point a. A similar statement holds as regards an interval on the left of a; and it is clear that the theorem can be stated so as to apply to the more general case of a neighbourhood which contains a in its interior.

It has been shewn in § 359, that the function $f(x)$ is differentiable within the interval $(a, a+R)$ of x, and is obtained by means of a term by term differentiation of the series. The same theorem may be applied to the function $f'(x)$, and to the series which represents it, and thus successively to the higher derivatives of $f(x)$. We have therefore

$$f^{(p)}(x) = \sum_{n=p}^{\infty} n(n-1)\dots(n-p+1)\, c_n (x-a)^{n-p};$$

* See *Math. Annalen*, vol. XLIV, "Ueber die nothwendigen und hinreichenden Bedingungen des Taylor'schen Lehrsatzes für Functionen einer reelen Variabeln."

hence $f(a) = c_0$, $f^{(p)}(a) = p!\,c_p$, where $f^{(m)}(a)$ is the derivative at a on the right; and thus the conditions (1) and (2) are satisfied.

We have now

$$f(a+h) = \sum_0^\infty \frac{1}{n!} f^{(n)}(a) \,.\, h^n,$$

$$f^{(p)}(a+h) = \sum_p^\infty \frac{1}{(n-p)!} f^{(n)}(a) \,.\, h^{n-p},$$

where $0 \leqq h < R$.

From the theorem of § 355, that a power-series converges absolutely at all points within its interval of convergence, we see that the function $\phi(x)$, defined, for the interval $a \leqq x < a + R$, by

$$\phi(x) = \sum_0^\infty |c_n| (x-a)^n,$$

has properties similar to those of $f(x)$; and thus that

$$\phi(a+h) = \sum_0^\infty |c_n| h^n = \sum_0^\infty \frac{1}{n!} \phi^{(n)}(a) \,.\, h^n,$$

$$\phi^{(p)}(a+h) = \sum_p^\infty \frac{n!}{(n-p)!} |c_n| h^{n-p} = \sum_p^\infty \frac{1}{(n-p)!} \phi^{(n)}(a) \,.\, h^{n-p},$$

for $0 \leqq h < R$. The functions $\phi(a+h)$, $\phi^{(p)}(a+h)$ are continuous functions of h in the interval $0 \leqq h < R$; and for each value of p,

$$|f^{(p)}(a+h)| \leqq \phi^{(p)}(a+h).$$

In order to prove that the condition (3) is satisfied, it will be sufficient to shew that the corresponding condition is satisfied for the function $\phi^{(n)}(a+h)$.

If $0 \leqq h \leqq h + k < R$, we have

$$\phi(a+h+k) = \sum_0^\infty \frac{1}{n!} \phi^{(n)}(a)(h+k)^n = \sum_0^\infty \frac{1}{n!} \phi^{(n)}(a+h) k^n;$$

and it will now be shewn that the series $\Sigma \frac{1}{n!} \phi^{(n)}(a+h) k^n$ converges uniformly for all values of h and k which are such that $0 \leqq h \leqq h + k \leqq r$, where $r < R$.

Let $\phi(a+h+k) = S_n(h, k) + T_n(h, k)$, where S_n denotes the sum of the first n terms of the series $\Sigma \frac{1}{n!} \phi^{(n)}(a+h) k^n$, and T_n denotes the remainder after those terms. Let h, k have arbitrarily assigned values which satisfy the condition above stated, and let δ be a positive number less than h and k, and such that $h + k + 2\delta < R$; also let ζ, η be two numbers in the interval $(-1, 1)$, arbitrarily chosen. In case $h = 0$, $k > 0$, the positive number δ is to be so chosen that $k + 2\delta < R$, and that $\delta < k$; also ζ is to be in the interval $(0, 1)$. Similarly if $h > 0$, $k = 0$, the number δ is to satisfy the conditions $h + 2\delta < R$, $\delta < h$; and η is to be in the interval $(0, 1)$. If $h = 0$, $k = 0$, then δ is to be so chosen that $2\delta < R$, and both ζ and η are to be in the interval (0, 1). We have then

$$\phi(a+h+k+\overline{\zeta+\eta}\delta) = S_n(h+\zeta\delta, k+\eta\delta) + T_n(h+\zeta\delta, k+\eta\delta),$$

and hence

$$T_n(h+\zeta\delta, k+\eta\delta) \leqq |\phi(a+h+k+\overline{\zeta+\eta}\delta) - \phi(a+h+k)| + |S_n(h+\zeta\delta, k+\eta\delta) - S_n(h, k)| + T_n(h, k).$$

If ϵ be a prescribed positive number, n may be chosen such that, for the prescribed values of h and k, $T_n(h, k) < \frac{1}{3}\epsilon$. Since $\phi(a+h+k)$ is a continuous function of $h+k$, and $S_n(h, k)$ is a continuous function of the two variables h, k; it follows that δ can be so fixed that

$$|\phi(a+h+k+\overline{\zeta+\eta}\delta) - \phi(a+h+k)| < \tfrac{1}{3}\epsilon,$$

and also
$$|S_n(h+\zeta\delta, k+\eta\delta) - S_n(h, k)| < \tfrac{1}{3}\epsilon.$$

The numbers n, δ have now been so chosen that, whatever values ζ and η may have, subject to the restrictions already stated,

$$T_n(h+\zeta\delta, k+\eta\delta) < \epsilon;$$

moreover, since the series $\Sigma \frac{1}{n!}\phi^{(n)}(a+h)k^n$ contains only positive terms, we have

$$T_\nu(h+\zeta\delta, k+\eta\delta) < \epsilon,$$

for $\nu \geqq n$.

It has thus been shewn that the series representing the function $\phi(a+h+k)$, considered as a function of the two variables (h, k), converges uniformly for a certain neighbourhood of each single point (h, k). Since there are no points in the neighbourhood of which $\phi(a+h+k)$ converges non-uniformly, it follows from the theorem of § 342, that the function converges uniformly for all values of h, k such that $0 \leqq h \leqq h+k \leqq r$, where $r < R$. Thus $T_\nu(h, k) < \epsilon$, for $\nu \geqq n$, and for all values of h, k which are not negative, and are such that $h+k \leqq r$.

The analogous result holds *a fortiori* for the series

$$\sum_0^\infty \frac{1}{(n+p)!}\phi^{(n)}(a+h).k^{n+p},$$

and also for the series

$$\sum_p^\infty \frac{1}{(n-p)!}\phi^{(n)}(a+h).k^{n-p},$$

which is obtained from $\Sigma \frac{1}{n!}\phi^{(n)}(a+h)k^n$ by differentiating p times with respect to k.

Since all these series contain positive terms only, it follows that n can be so chosen that

$$\frac{1}{(\nu \pm p)!}\phi^{(\nu)}(a+h)\,.\,k^{\nu \pm p} < \epsilon$$

for $\nu \geqq n$, and $0 \leqq h \leqq h+k \leqq r$; and thus the condition (3) in the statement of the theorem is satisfied, since $|f^{(\nu)}(a+h)| \leqq \phi^{(\nu)}(a+h)$.

366. *If $f^{(n)}(x)$ be defined for every integral value of n, where x is such that $a \leqq x < a+R$, and if, for some one value of p which is a positive or negative integer, or zero, it satisfy the condition that $\frac{1}{(n+p)!}f^{(n)}(a+h)k^{n+p}$ converges to zero when n is indefinitely increased, uniformly for all values of h, k, such that $0 \leqq h \leqq h+k \leqq r$, where $r < R$, then the same condition holds for every value of p which is integral or zero.*

If we denote $\frac{1}{(n+p)!}f^{(n)}(a+h)k^{n+p}$ by $F_{p,n}(h,k)$, we have

$$F_{p+1,n}(h,k) = \frac{k}{n+p+1}F_{p,n}(h,k);$$

and hence, since $k \leqq r$, $|F_{p+1,n}(h,k)| < |F_{p,n}(h,k)|$, if $n+p+1 > r$. It follows that $\lim_{n=\infty} F_{p+1,n}(h,k)$, and generally $\lim_{n=\infty} F_{p+q,n}(h,k)$, for $q > 0$, converges uniformly if $F_{p,n}(h,k)$ does so.

Again
$$F_{p-1,n}(h,k) = \frac{1}{(n+p-1)!}f^{(n)}(a+h)k^{n+p-1}$$
$$= F_{p,n}(h,k+\delta)\left(\frac{k}{k+\delta}\right)^{n+p-1}\frac{n+p}{k+\delta};$$

if $r < R$ be fixed, δ can be so chosen as to be positive, and such that $r+\delta < R$. Hence, if $0 \leqq h \leqq r$,

$$|F_{p-1,n}(h,k)| < |F_{p,n}(h,k+\delta)|\left(\frac{r}{r+\delta}\right)^{n+p-1}\frac{n+p}{\delta};$$

if n be so chosen that

$$\left(\frac{r}{r+\delta}\right)^{n+p-1}\frac{n+p}{\delta} \leqq 1, \text{ and } |F_{p,n}(h,k+\delta)| < \epsilon,$$

for $0 \leqq h \leqq h+k \leqq r$, we have, then, for such values of n, $|F_{p-1,n}(h,k)| < \epsilon$. It is now seen, also, that the corresponding result holds for $|F_{p-q,n}(h,k)|$, where $q > 0$.

367. The necessary and sufficient conditions that Taylor's theorem should hold for the function $f(a+h)$, where $0 \leqq h < R$, can be most simply expressed if Cauchy's form of the remainder be used, and may be obtained as follows:—

The condition as to the existence of differential coefficients of all orders being assumed to be satisfied, it has now been shewn in § 365, to be a necessary condition for the validity of Taylor's series, that $\frac{1}{(n-1)!}f^{(n)}(a+h)\,k^{n-1}$ should converge to the limit zero, as n is indefinitely increased, uniformly for all values of h and k such that $0 \leqq h \leqq h+k \leqq r$, where $r < R$. If we write θh for h, and $(1-\theta)h$ for k, then, if the condition be satisfied, the expression

$$\frac{1}{(n-1)!}f^{(n)}(a+\theta h)(1-\theta)^{n-1}h^{n-1}$$

converges to zero, when n is indefinitely increased, uniformly for all values of θ and h such that $0 \leqq \theta \leqq 1$, and $0 \leqq h \leqq r$; and it follows that, for each value of h, $\frac{1}{(n-1)!}f^{(n)}(a+\theta h)(1-\theta)^{n-1}h^n$ converges to zero uniformly for all values of θ. It has been shewn in § 364, that this last condition is sufficient to secure the convergence of the series to the sum $f(a+h)$. We have now established the following theorem:—

That the function $f(a+h)$, defined for all values of h such that $0 \leqq h < R$, may be represented for all the values of h by the series $\sum\limits_0^\infty \frac{1}{n!}f^{(n)}(a)h^n$, it is necessary and sufficient, (1) *that $f(x)$ have differential coefficients of all orders for $a < x < a+R$, and derivatives at the point a on the right, of all orders; and* (2) *that Cauchy's remainder $\frac{1}{(n-1)!}f^{(n)}(a+\theta h)(1-\theta)^{n-1}h^n$, for each value of h such that $0 \leqq h < R$, converge to zero, when n is indefinitely increased, uniformly for all values of θ in the interval* (0, 1).

In case Lagrange's form of the remainder in Taylor's theorem be employed instead of that due to Cauchy, the necessary and sufficient conditions cannot be stated in so simple a form. The following theorem has reference to this form of the remainder:—

In order that the function $f(a+h)$, defined for all values of h such that $0 \leqq h < R$, may be represented, for all the values of h, by the series $\Sigma \frac{h^n}{n!}f^{(n)}(a)$, it is necessary, besides the condition of unrestricted differentiability previously stated, that $\frac{1}{n!}f^{(n)}(a+\theta h)h^n$ should converge, for each value of h such that $0 \leqq h < \frac{1}{2}R$, to the limit zero, when h is indefinitely increased, uniformly for all values of θ in the interval (0, 1). *It is not necessary, but it is sufficient, that the expression should converge to zero for each value of h such that $0 \leqq h < R$, uniformly for all values of θ in* (0, 1).

In accordance with the theorem proved in § 365, it is necessary that $\frac{1}{n!}f^{(n)}(a+h)k^n$ should converge to the limit zero, when n is indefinitely increased, uniformly for all values (h, k) which are such that $0 \leqq h \leqq h+k \leqq r$, where $r < R$. If we write h for k, and θh for h, we see that this condition includes the condition that $\frac{1}{n!}f^{(n)}(a+\theta h)h^n$ should converge to zero, for each value of h such that $0 \leqq h < \frac{1}{2}R$, uniformly for all values of θ in the interval $(0, 1)$. This condition is therefore a necessary one.

Let us next assume that $\frac{1}{n!}f^{(n)}(a+\theta h)h^n$ converges to zero, for each fixed value of h such that $0 \leqq h < R$, uniformly for all values of θ in the interval $(0, 1)$. Let ρ, r be two positive numbers such that $\rho \leqq r < R$, and in the expression $\frac{1}{n!}f^{(n)}(a+\theta h)h^n$ write $h = r$, $\theta h = \rho$; then $\lim\limits_{n=\infty} \frac{1}{n!}f^{(n)}(a+\rho)r^n = 0$. It follows that the series $\sum\limits_0^\infty \frac{1}{n!}f^{(n)}(a+\rho)k^n$ converges absolutely for every positive value of k which is $< r$, and therefore for every positive value of k which is $< R$, since r may be taken arbitrarily near to R. In the series $\Sigma \frac{1}{n!}f^{(n)}(a+\rho)k^n$, the functions $f^{(n)}(a+\rho)$ may be replaced by the absolutely convergent series in powers of ρ; and by the theorem of § 358, the terms of the series $\Sigma \frac{1}{n!}f^{(n)}(a+\rho)k^n$, thus expressed, may be rearranged without affecting the convergence or the sum of the series. The series then becomes $\Sigma \frac{1}{n!}f^{(n)}(a)(\rho+k)^n$, and this series converges for $\rho < R$, $k < R$, *i.e.* for $\rho + k < 2R$. It follows that the series $\Sigma \frac{1}{n!}f^{(n)}(a)h^n$ converges if $h < 2R$. Now it is clearly not necessary for the validity of Taylor's theorem within the range $0 \leqq h < R$, that the series should converge for all values of h which are $< 2R$; for R might be the upper limit of the values of h for which the power-series converges, and then the series could not converge for values of h which are $> R$. It has thus been shewn that the condition for all values of h in the interval $0 \leqq h < R$, is not a necessary one. It is clearly however a sufficient condition.

It was remarked by Cauchy* that the series $\Sigma \frac{h^n}{n!}f^{(n)}(a)$ may be convergent in an interval, and yet that its sum need not be $f(a+h)$. This happens whenever the remainder R_n, in Taylor's theorem, which is defined as the difference between $f(a+h)$ and the sum of the first n terms of the series, converges, for each value of x, to a limit which is different from zero, as n is indefinitely increased.

Let the function $f(x)$ be defined by $f(x) = e^{-\frac{1}{x^2}}$, for $x^2 > 0$, and $f(0) = 0$; it can easily be shewn that this function and all its differential coefficients

* *Calcul Diff.* p. 108; see also P. Du Bois Reymond, *Math. Annalen*, vol. XXI, p. 114.

exist and are zero at the point $x=0$; and that for $x^2>0$, the remainder in the Taylor's series has for its limit $e^{-\frac{1}{x^2}}$. If now $\phi(x)=e^x+e^{-\frac{1}{x^2}}$, $(x^2>0)$, $\phi(0)=1$, and the series $\Sigma \frac{h^n}{n!}\phi^{(n)}(0)$ in the neighbourhood of the point $x=0$ be formed, then the series converges, not to the value $\phi(h)$, but to the value e^h.

EXAMPLES.

1. Let $f(x)=(1+x)^p$; then, in a neighbourhood of the point $x=0$, we have

$$f(x)=1+px+\frac{p(p-1)}{2!}x^2+\ldots+\frac{p(p-1)\ldots(p-n+2)}{(n-1)!}x^{n-1}+R_n,$$

where R_n can be expressed in Lagrange's form by

$$\frac{p(p-1)\ldots(p-n+1)}{n!}\frac{x^n}{(1+\theta x)^{n-p}},$$

or in Cauchy's form by

$$\frac{p(p-1)\ldots(p-n+1)}{(n-1)!}\frac{(1-\theta)^{n-1}}{(1+\theta x)^{n-p}}x^n.$$

Using Cauchy's form, we see that

$$|R_n| \leqq \left|\frac{p(p-1)\ldots(p-n+1)}{(n-1)!}x^n\right|,$$

provided $n>p$. If $|x|<1$, the expression

$$\left|\frac{p(p-1)\ldots(p-n+1)}{(n-1)!}x^n\right|$$

continually diminishes as n is increased: for, denoting it by u_n, we find

$$\frac{u_{n+1}}{u_n}=\left|\frac{p-n}{n}x\right|<1-\epsilon,$$

where ϵ is a fixed positive number $<1-|x|$, provided n be sufficiently great, and it follows that the limit of u_n is zero; and thus $\lim R_n=0$. The series therefore converges for all values of x such that $|x|<1$.

To find the limit of $\left|\frac{p(p-1)\ldots(p-n+1)}{n!}\right|$ when n is indefinitely increased, suppose first that $p+1$ is negative, say $-k$. We may write the expression in the form

$$\left(1+\frac{k}{1}\right)\left(1+\frac{k}{2}\right)\ldots\left(1+\frac{k}{n}\right),$$

and this is $>1+k\left(\frac{1}{1}+\frac{1}{2}+\ldots+\frac{1}{n}\right)$; thus the limit is indefinitely great. Next suppose that $p+1$ is positive. Then the expression may be written in the form

$$\frac{p(p-1)\ldots(p-\lambda+2)}{(\lambda-1)!}\left(1-\frac{p+1}{\lambda}\right)\left(1-\frac{p+1}{\lambda+1}\right)\left(1-\frac{p+1}{\lambda+2}\right)\ldots\left(1-\frac{p+1}{n}\right)$$

where λ is the integer next greater than $p+1$; this is less than

$$\frac{p(p-1)\ldots(p-\lambda+2)}{(\lambda-1)!}\frac{1}{\left(1+\frac{p+1}{\lambda}\right)\left(1+\frac{p+1}{\lambda+1}\right)\ldots\left(1+\frac{p+1}{n}\right)},$$

or than $$\frac{p(p-1)\ldots(p-\lambda+1)}{(\lambda-1)!}\frac{1}{1+(p+1)\left(\frac{1}{\lambda}+\frac{1}{\lambda+1}+\ldots\frac{1}{n}\right)};$$

hence the limit, when n is indefinitely increased, is zero. If $p=-1$, the limit is unity.

If $x=1$, Lagrange's form of the remainder shews that the series converges if $p>-1$. The series diverges if $p<-1$, because the general term of the series increases indefinitely with n. The series oscillates if $p=-1$.

If $x=-1$, Cauchy's form of the remainder shews that if $p-1>-1$, or $p>0$, the series is convergent. It is divergent if $p<0$, for the sum of n terms of the series is

$$(-1)^{n-1}\frac{(p-1)(p-2)\ldots(p-n+1)}{(n-1)!}.$$

2. Let* $f(x)=\sum_{r=0}^{\infty}\frac{(-1)^r}{r!}\frac{a^{-r}}{a^{-2r}+x^2}$, where $a>1$. For this function

$$f(0)=e^{-a},\quad f^{(2k-1)}(0)=0,\quad f^{(2k)}(0)=(-1)^k(2k)!\,e^{-a^{2k+1}};$$

thus the series for $f(x)$ is

$$\sum_0^{\infty}(-1)^k e^{-a^{2k+1}}x^{2k},$$

which is everywhere convergent.

The sum of the series, for $x=0$, is $f(0)$, but in every neighbourhood of $x=0$, the sum of the series and the value of $f(x)$ are different except at most at a finite number of points.

3. Let $f(x)=\sum_{r=0}^{\infty}\frac{1}{r!}\frac{a^{-r}}{a^{-2r}+x^2}$, where $a>1$. For this function, the Maclaurin's series is $\Sigma(-1)^k e^{a^{2k+1}}x^{2k}$, which diverges for every value of x except $x=0$.

4. Let† $f(x)=\sum_0^{\infty}\frac{(-1)^n}{n!}\frac{1}{1+a^n x}$, where $a>1$. This function is continuous on the right of the point $x=0$, and has derivatives on the right of all orders at that point; the Maclaurin's series $\Sigma(-1)^n\left(\frac{1}{e}\right)^{a^n}.x^n$ thus obtained, converges for all positive values of x, but does not represent the function $f(x)$.

5. Let $f(x)=\sum_0^{\infty}\frac{1}{n!}\frac{1}{1+a^n x}$, where $a>1$. For this function the Maclaurin's series does not converge in any neighbourhood of the point $x=0$.

MAXIMA AND MINIMA OF A FUNCTION OF ONE VARIABLE.

368. It has been shewn in § 207, to be a necessary condition that a function $f(x)$ may have an extreme at the point $x=0$, that the differential coefficient at that point should be zero, provided the function be such that a differential coefficient at $x=0$ exists. Let us assume the function to be such that the first n differential coefficients $f'(x), f''(x), \ldots f^{(n)}(x)$ all exist and are continuous, at every point x such that $-\delta<x<\delta$. Let us further assume that $f'(0), f''(0), \ldots f^{(n-1)}(0)$ are all zero, and thus that $f^{(n)}(0)$ is that differential coefficient of lowest order which does not vanish at $x=0$.

* Pringsheim, *Münchener Berichte*, 1892, p. 222.

† Pringsheim, *Math. Annalen*, vol. XLII, p. 161, and vol. XLIV, p. 54.

We have then $f(x) - f(0) = \frac{x^n}{n!} f^{(n)}(\theta x)$; where $0 < \theta < 1$, and x is such that $-\delta < x < \delta$. Since $f^{(n)}(x)$ is continuous at $x = 0$, a neighbourhood $(-\delta', \delta')$ of that point, interior to $(-\delta, \delta)$, can be so determined, that $f^{(n)}(\theta x)$ has the same sign as $f^{(n)}(0)$, provided $-\delta' \leqq x \leqq \delta'$. If n be odd, the sign of $f(x) - f(0)$, in the interval $(-\delta', \delta')$, depends upon that of x; and therefore $f(x)$ has neither a maximum nor a minimum at the point $x = 0$. If n be even, the sign of $f(x) - f(0)$ is the same as that of $f'(0)$, in the whole interval $(-\delta', \delta')$, and therefore $f(x)$ has a maximum or a minimum at $x = 0$, according as $f^{(n)}(0)$ is negative or positive. The following theorem for determining whether a maximum, or a minimum exists at a point at which the differential coefficient of a function $f(x)$ vanishes has therefore been established:—

If the first n differential coefficients of a function $f(x)$ all exist, and are continuous, at all interior points of the interval $(-\delta, \delta)$; and if $f^{(n)}(x)$ be the differential coefficient of lowest order which does not vanish at the point $x = 0$, then (1) if n be odd, there is neither a maximum nor a minimum of the function $f(x)$ at the point $x = 0$; and (2) if n be even, the point $x = 0$ is a maximum or a minimum of $f(x)$, according as $f^{(n)}(0)$ is negative or positive.

It is unnecessary for the application of the criterion given in this theorem that $f(x)$ be capable of representation in a neighbourhood of the point $x = 0$ by a convergent power-series. The theorem cannot be applied to the case of a function with differential coefficients of all orders, when they all vanish at the point $x = 0$.

EXAMPLES.

1.* Let $f(x) = x^2 - e^{-\frac{1}{x^2}}$, and $f(0) = 0$. In this case $f'(0) = 0$, $f''(0) = 2$; and $f'(x)$, $f''(x)$ are continuous in any neighbourhood of $x = 0$. The theorem establishes that $f(x)$ has a minimum at $x = 0$, although $f(x)$ cannot be represented by a power-series in any neighbourhood of the point.

2.* The function defined by $f(x) = e^{-\frac{1}{x^2}}$, $f(0) = 0$, has a minimum at $x = 0$; and yet the theorem is not applicable, because the differential coefficients of all orders vanish at $x = 0$.

3.* The function defined by $f(x) = xe^{-\frac{1}{x^2}}$, $f(0) = 0$, has neither a maximum nor a minimum at $x = 0$. As in (2), the above theorem is in this case inapplicable.

* These examples are given by Scheeffer, *Math. Annalen*, vol. xxxv, p. 542.

TAYLOR'S THEOREM FOR FUNCTIONS OF TWO VARIABLES.

369. Let us assume a function $f(x, y)$ to be defined for all values of x, y in the domain defined by $a-\delta \leqq x \leqq a+\delta$, $b-\delta' \leqq y \leqq b+\delta'$. Under proper conditions as to the existence and continuity of the partial differential coefficients of $f(x, y)$, of a finite number n of orders, it is possible to obtain an expression for $f(a+h, b+k)-f(a, b)$ consisting of terms involving the first n powers of h and k, together with a remainder analogous to the remainder in Taylor's theorem, such expression being valid for values of h, k, such that $|h|<\delta$, $|k|<\delta'$. It is however, for the present purpose, unnecessary to consider the least stringent set of conditions relating to the partial differential coefficients of the various orders, which are sufficient to allow the extension of Taylor's theorem to the case of a function of two variables. It will here be assumed that, for all values of x and y such that $a-\delta<x<a+\delta$, $b-\delta'<y<b+\delta'$, the partial differential coefficients of $f(x, y)$ of the first n orders all exist, and are finite; and further, that they are all continuous, for this range of values of x and y, with respect to (x, y). In accordance with the theorem of § 240, the order of differentiation, in each of the mixed partial differential coefficients, is in this case immaterial.

Taking values of h and k which are numerically less than δ, δ' respectively, let $f(a+th, b+tk)$ be denoted by $F(t)$, the variable t having the domain $(-1, +1)$. The conditions contained in the last theorem of § 236 being in this case satisfied, the differential coefficient $F'(t)$ of $F(t)$ exists, and is equal to $\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)f(x, y)$, where $x=a+th$, $y=b+tk$. Similarly, it is seen that all the differential coefficients

$$F''(t), F'''(t), \dots F^{(n)}(t)$$

exist and are continuous; and that

$$F^{(r)}(t)=\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^r f(x, y).$$

In accordance with the theorem of § 363, we have

$$F(t)=F(0)+tF'(0)+\frac{t^2}{2!}F''(0)+\dots+\frac{t^{n-1}}{(n-1)!}F^{(n-1)}(0)+\frac{t^n}{n!}F^{(n)}(\theta t),$$

where θ is a number such that $0<\theta<1$.

Since this holds for $t=1$, we see that

$$f(a+h, b+k)=f(a, b)+\left(h\frac{\partial}{\partial a}+k\frac{\partial}{\partial b}\right)f(a, b)+\frac{1}{2!}\left(h\frac{\partial}{\partial a}+k\frac{\partial}{\partial b}\right)^2 f(a, b)+\dots$$

$$\dots+\frac{1}{(n-1)!}\left(h\frac{\partial}{\partial a}+k\frac{\partial}{\partial b}\right)^{n-1}f(a, b)+\frac{1}{n!}\left(h\frac{\partial}{\partial a}+k\frac{\partial}{\partial b}\right)^n f(a+\theta h, b+\theta k).$$

This is an extension of Taylor's theorem to the case of a function of two variables. It has been established for all values of h, k such that $|h| < \delta$, $|k| < \delta'$, on the hypothesis that $f(x, y)$ and all its partial differential coefficients exist for all values of x and y such that $a - \delta < x < a + \delta$, $b - \delta' < y < b + \delta'$, and that they are all continuous with respect to the two-dimensional continuum (x, y).

MAXIMA AND MINIMA OF FUNCTIONS OF TWO VARIABLES.

370. Necessary and sufficient conditions have been stated, in the theorem of § 242, that the point (0, 0) may be a point at which a function $f(x, y)$ has a maximum, or a minimum. The general theory of maxima and minima of functions of two variables has been discussed by Scheeffer*, Dantscher†, and Stolz‡, the last of whom has extended Scheeffer's method to the case of functions of any number of variables. The account which will here be given of the general theory is based upon the investigations of Scheeffer, as modified by Stolz.

Let the function $f(x, y)$ be such that either $f(x, y) - f(0, 0)$ is representable in a neighbourhood of the point (0, 0) by a convergent series consisting of powers of x and y, or else that it is such that the theorem of § 369 is applicable, so that

$$f(x, y) - f(0, 0) = G_n(x, y) + R_{n+1}(x, y),$$

where $G_n(x, y)$ consists of terms of dimensions not higher than n, in x and y; and $R_{n+1}(x, y)$ is either a convergent series of which the terms of lowest dimension are of the order $n + 1$, or has the form of the remainder given in § 369, consisting of terms of dimension $n + 1$ in θx, θy, where $0 < \theta < 1$; and in the latter case it will be assumed that the differential coefficients in that remainder are limited in the whole domain. It will be shewn that, under a certain condition, the problem of determining whether the point (0, 0) is a point at which $f(x, y)$ has a maximum or a minimum is reducible to the solution of the corresponding problem relating to the rational integral function $G_n(x, y)$. The following general theorem will be established:—

The function $f(x, y)$ *having in the neighbourhood of* (0, 0) *the character above described, if an index* n *and two positive numbers* c, δ *can be so determined that* (1) *for all values of* x *such that* $0 < |x| < \delta$, *the upper and lower limits of* $G_n(x, y)$, *for a constant value of* x, *and for all values of* y *in the interval* $(-x, x)$, *are in absolute value not less than* $c|x|^n$; *and* (2) *that, if* $0 < |y| < \delta$, *the upper and lower limits of* $G_n(x, y)$, *for a constant*

* *Math. Annalen*, vol. XXXV.

† *Math. Annalen*, vol. XLII.

‡ *Berichte* of the Vienna Academy, vols. XCIX, C; also *Grundzüge*, vol. I, p. 211.

value of y, and for all values of x in the interval $(-y, y)$, are in absolute value not less than $c\,|y|^n$; then the two functions $f(x, y)$, $G_n(x, y)$ have both either a proper maximum, or both a proper minimum, or both neither a maximum nor a minimum, at the point $(0, 0)$.

To prove this theorem, we first observe that $R_{n+1}(x, y)$ can be regarded as a homogeneous function of x and y of degree $n+1$, in which the coefficients depend upon x and y. By giving each of the coefficients its greatest possible value, for $|x| < \delta$, $|y| < \delta$, we see that

$$|R_{n+1}(x, y)| < A_0|x|^{n+1} + A_1|x|^n|y| + \ldots + A_{n+1}|y|^{n+1};$$

where $A_0, A_1, \ldots A_{n+1}$ are positive numbers.

If now $|y| \leqq |x|$, we have

$$|R_{n+1}(x, y)| < (A_0 + A_1 + \ldots + A_{n+1})\,|x|\,\,|x|^n;$$

hence we see that a number $\delta' < \delta$ can be so chosen that

$$|R_{n+1}(x, y)| < \epsilon\,|x|^n,$$

where ϵ is an arbitrarily chosen positive number, provided $|x| < \delta'$, $|y| \leqq |x|$. In a similar manner we can shew that δ' can be so chosen that

$$|R_{n+1}(x, y)| < \epsilon\,|y|^n,$$

provided $|x| \leqq |y|$, and $|y| < \delta'$.

Let now the upper and the lower limits of $G_n(x, y)$, for a constant value of x, and for all values of y such that $|y| \leqq |x|$, be denoted by $G_n(x, \bar{\phi}(x))$, $G_n(x, \underline{\phi}(x))$ respectively. Also let the upper and the lower limits of $G_n(x, y)$, for a constant value of y, and for all values of x such that $|x| \leqq |y|$, be denoted by $G_n(\bar{\psi}(y), y)$, $G_n(\underline{\psi}(y), y)$ respectively. We have then, provided $|x| < \delta'$, and $|y| \leqq |x|$,

$$G_n(x, \underline{\phi}(x)) - \epsilon\,|x|^n < f(x, y) - f(0, 0) < G_n(x, \bar{\phi}(x)) + \epsilon\,|x|^n;$$

also, provided $|y| < \delta'$, $|x| \leqq |y|$, we have

$$G_n(\underline{\psi}(y), y) - \epsilon\,|y|^n < f(x, y) - f(0, 0) < G_n(\bar{\psi}(y), y) + \epsilon\,|y|^n.$$

First, let us assume that $G_n(0, 0)$ is a proper minimum of $G_n(x, y)$, and that the conditions of the theorem are satisfied. By the theorem of § 242, $G_n(x, \underline{\phi}(x))$, $G_n(\underline{\psi}(y), y)$ are both positive, for sufficiently small values of x and y; we may suppose δ' to be so small that these conditions are satisfied, provided $|x| < \delta'$, $|y| < \delta'$.

We have then $G_n(x, \underline{\phi}(x)) \geqq c\,|x|^n$, if $0 < |x| < \delta'$, $|y| \leqq |x|$; and $G_n(\underline{\psi}(y), y) \geqq c\,|y|^n$, if $0 < y < \delta'$, $|x| \leqq |y|$.

It now follows that

$$(c-\epsilon)\,|x|^n < f(x, y) - f(0, 0), \text{ for } 0 < |x| < \delta', \ |y| \leqq |x|,$$

and that

$$(c-\epsilon)|y|^n < f(x, y) - f(0, 0), \text{ for } 0<|y|<\delta', \ |x| \leqq |y|.$$

Since ϵ can be chosen so as to be less than c, we see that $f(x, y)-f(0, 0)$ is positive for all values of x and y such that $0<|x|<\delta'$, $0<|y|<\delta'$, and therefore $f(0, 0)$ is a proper minimum of $f(x, y)$.

Next, let us assume that $G_n(0, 0)$ is a proper maximum of $G_n(x, y)$; then $G_n(x, \bar{\phi}(x))$, $G_n(\bar{\psi}(y), y)$ are both negative, for sufficiently small values of x and y. We therefore assume that

$$G_n(x, \bar{\phi}(x)) \leqq -c|x|^n, \text{ for } 0<|x|<\delta', \text{ and } |y| \leqq |x|;$$

and that

$$G_n(\bar{\psi}(y), y) \leqq -c|y|^n, \text{ for } 0<|y|<\delta', \ |x| \leqq |y|.$$

We have then $f(x, y)-f(0, 0) < -(c-\epsilon)|x|^n$, for $0<|x|<\delta'$, and $|y| \leqq |x|$; and also $f(x, y)-f(0, 0) < -(c-\epsilon)|y|^n$, for $0<|y|<\delta'$, $|x| \leqq |y|$. Since ϵ may be taken to be $<c$, it follows that $f(0, 0)$ is a proper maximum of $f(x, y)$.

Lastly, let us assume that $G_n(0, 0)$ is neither a maximum nor a minimum of $G_n(x, y)$. In this case we may, for example, assume that

$$G_n(x, \bar{\phi}(x)) \geqq cx^n, \ G_n(x, \underline{\phi}(x)) \leqq -cx^n, \text{ for } 0<x<\delta.$$

We have then, $f(x, \bar{\phi}(x)) - f(0, 0) > (c-\epsilon)x^n$,

and $f(x, \underline{\phi}(x)) - f(0, 0) < -(c-\epsilon)x^n$,

provided $0<x<\delta'$. Since ϵ may be taken to be less than c, these two differences are of opposite signs; therefore $f(0, 0)$ is neither a maximum nor a minimum of $f(x, y)$.

It should be observed that this theorem does not always suffice to decide whether the point $(0, 0)$ is a point at which $f(x, y)$ has an extreme value, or not. For it may happen that, for a given function $f(x, y)$ of the assumed type, no value of n can be determined, for which the conditions stated in the theorem hold, and therefore the theorem is inapplicable however great n may be taken.

If $f(x, y) = [u(x, y)]^2$, where $u(x, y)$ vanishes at points of a locus which passes through the point $(0, 0)$, then the function $f(x, y)$ is one for which the theorem is inapplicable; the point $(0, 0)$ is in this case a point at which $f(x, y)$ has an improper minimum.

In general the theorem is inapplicable in the case of any function which attains the value zero, at points other than $(0, 0)$, in every neighbourhood of that point, but which has one and the same sign at all points at which it does not vanish.

371. The simplest case in which the theorem of § 370 can be applied is that in which the function $G_n(x, y)$ is a homogeneous function of degree n. For such a function $G_n(x, y)$, three cases arise.

(1) If $G_n(x, y)$ be a definite form, *i.e.* if $G_n(x, y)$ has one and the same sign for all values of (x, y) except $(0, 0)$, then $G_n(0, 0)$ is a proper minimum, or a proper maximum, according as that sign is positive or negative.

(2) If $G_n(x, y)$ be an indefinite form, *i.e.* if there are points in every neighbourhood of $(0, 0)$ at which $G_n(x, y)$ is positive, and others at which it is negative, there are other points besides $(0, 0)$ at which the function vanishes, and there is no extreme of the function $G_n(x, y)$ at the point $(0, 0)$.

(3) If $G_n(x, y)$ be semi-definite, *i.e.* if $G_n(x, y)$ vanishes at points other than $(0, 0)$, but has a fixed sign at all points at which it does not vanish, then $G_n(0, 0)$ is an improper extreme of $G_n(x, y)$.

It should be observed that, if n be odd, $G_n(x, y)$ is necessarily an indefinite form.

It will be shewn that, when $G_n(x, y)$ is definite or indefinite, it satisfies the conditions stated in the theorem of § 370; accordingly $f(x, y)$ has a proper maximum or else a proper minimum, when $G_n(x, y)$ is a definite form; and $f(x, y)$ has no extreme when $G_n(x, y)$ is an indefinite form.

When $G_n(x, y)$ is a semi-definite form, no conclusion can be drawn as to the existence of an extreme of the function $f(x, y)$, as the conditions contained in the theorem of § 370 are not satisfied.

If $G_n(x, y)$ be definite, it is of the form

$$G_n(x, y) = A \prod_{r=1}^{r=k} \{(y - \gamma_r x)^2 + \delta_r^2 x^2\},$$

where $n = 2k$. Let us assume that A is positive; then

$$G_n(x, y) \geqq A \prod_{r=1}^{r=k} \delta_r^2 . x^n,$$

for all values of x and y; it follows that the first condition of the theorem is satisfied.

The case in which A is negative may be treated in a similar manner.

Again

$$(y - \gamma_r x)^2 + \delta_r^2 x^2 = \left(x\sqrt{\gamma_r^2 + \delta_r^2} - \frac{\gamma_r y}{\sqrt{\gamma_r^2 + \delta_r^2}}\right)^2 + \frac{\delta_r^2}{\gamma_r^2 + \delta_r^2} y^2 > \frac{\delta_r^2}{\gamma_r^2 + \delta_r^2} y^2,$$

hence $$|G_n(\bar{\psi}(y), y)| > |G_n(\underline{\psi}(y), y)| \geqq |A| y^n \prod_1^k \frac{\delta_r^2}{\gamma_r^2 + \delta_r^2};$$

and therefore the second condition of the theorem is satisfied.

Next let $G_n(x, y)$ be an indefinite form; in which case $G_n(x, y)$ has neither a maximum nor a minimum at $(0, 0)$. Let (x', y') be a point at which $G_n(x', y') > 0$; and first suppose that $|y'| \leqq |x'|$, so that $|x'| > 0$.

Let x, y be such that $y/x = y'/x'$, and let x, x' have the same sign; we have $G_n(x, y) > 0$, and it follows that

$$G_n(x, \bar{\phi}(x)) \geqq \frac{G_n(x', y')}{|x'|^n} |x|^n > 0.$$

Next suppose that $|x'| \leqq |y'|$, so that $|y'| > 0$; we then shew in the same manner that

$$G_n(\bar{\psi}(y), y) \geqq \frac{G_n(x', y')}{|y'|^n} |y|^n > 0,$$

where y has the same sign as y'.

Since there are also values of (x', y') such that $G_n(x', y') < 0$, we can shew as before that

$$G_n(x, \underline{\phi}(x)) \leqq \frac{G_n(x', y')}{|x'|^n} |x|^n < 0,$$

where x and x' have the same sign, and that

$$G_n(\underline{\psi}(y), y) \leqq \frac{G_n(x', y')}{|y'|^n} |y|^n < 0,$$

where y has the same sign as y'. It has thus been established that, when $G_n(x, y)$ is an indefinite form, the conditions of the theorem of § 370 are satisfied.

The following general result has now been obtained:—

If $f(x, y) - f(0, 0)$ be of the form $G_n(x, y) + R_{n+1}(x, y)$, where $G_n(x, y)$ is a homogeneous function of degree n, then, if n be odd, $f(0, 0)$ is not an extreme of $f(x, y)$. If n be even, and $G_n(x, y)$ be an indefinite form, $f(0, 0)$ is not an extreme of $f(x, y)$. If $G_n(x, y)$ be a definite form, $f(0, 0)$ is a proper minimum, or a proper maximum, of $f(x, y)$, according as $G_n(x, y)$ is positive or negative. If $G_n(x, y)$ be a semi-definite form, no conclusion can be drawn from the consideration of $G_n(x, y)$ by itself, as to the existence or non-existence of an extreme of $f(x, y)$ at the point $(0, 0)$.

372. When those terms in the expansion of $f(x, y)$ in powers of x and y, which are of the lowest degree, give a semi-definite form, it is necessary to take a value of n greater than this lowest degree; we have therefore to consider the case in which $G_n(x, y)$ is not homogeneous. We have then, in order to apply the theorem of § 242, to $G_n(x, y)$, to determine the four functions $G_n(x, \bar{\phi}(x))$, $G_n(x, \underline{\phi}(x))$, $G_n(\bar{\psi}(y), y)$, $G_n(\underline{\psi}(y), y)$. The values $y = \bar{\phi}(x)$, $y = \underline{\phi}(x)$, may be either in the interior, or at the ends of the interval $(-x, x)$. In the former case they must be such as to satisfy the condition $\frac{dG_n(x, y)}{dy} = 0$; in the latter case they will in general not satisfy this condition, although they may do so. The method of procedure, by which $G_n(x, \bar{\phi}(x))$, $G_n(x, \underline{\phi}(x))$ may be obtained, is to obtain the various solutions of the equation $\frac{dG_n(x, y)}{dy} = 0$, in which y is expressed as a series of fractional or integral powers of x; only such values of y need be considered, as vanish for $x = 0$.

Let $y = P_1(x)$, $y = P_2(x)$, ... $y = P_r(x)$ denote these series; we then form the expressions $G_n(x, -x)$, $G_n(x, x)$, $G_n(x, P_1(x))$, ... $G_n(x, P_r(x))$.

It is certain that the two expressions $G_n(x, \bar{\phi}(x))$, $G_n(x, \underline{\phi}(x))$ must both occur amongst these $r + 2$ expressions, and a comparison of the leading terms of these expressions will enable us to identify the two expressions required. If the indices of the leading terms in $G_n(x, \bar{\phi}(x))$, $G_n(x, \underline{\phi}(x))$, are not greater than n, the first condition of the general theorem is satisfied. A similar method, in which the equation $\frac{dG_n(x, y)}{dx} = 0$ is employed, will lead to the determination of $G_n(\bar{\psi}(y), y)$, $G_n(\underline{\psi}(y), y)$.

The details of the investigation have been fully carried out by Scheeffer, who employs the somewhat more symmetrical, but practically less simple, method, in which x and y are expressed as series involving a single parameter.

When, for any value of n, the result of this process is that $G_n(x, y)$ is such that the conditions contained in the theorem of § 370 are not satisfied, a larger value of n in which more terms of $f(x, y)$ are included in $G_n(x, y)$ must be taken, and the process repeated until a definite result is obtained.

EXAMPLES.

1. Let $f(x, y) - f(0, 0) = ax^2 + 2hxy + by^2 + R_3(x, y)$. The form $ax^2 + 2hxy + by^2$ is definite if $ab - h^2$ is positive; in this case $f(0, 0)$ is a proper minimum or a proper maximum of $f(x, y)$, according as a is positive or negative. If $ab - h^2$ is negative, then $ax^2 + 2hxy + by^2$ is an indefinite form, and in that case $f(0, 0)$ is not an extreme of $f(x, y)$. If $ab - h^2 = 0$, the form $ax^2 + 2hxy + by^2$ is semi-definite, and no conclusion can be drawn as to the existence of an extreme of $f(x, y)$. It will be necessary in the last case to consider terms of order higher than 2 as included in $G_n(x, y)$. By taking $n = 3, 4, \ldots$ a function $G_n(x, y)$ may be determinable which satisfies the conditions of the theorem of § 370.

2.* Let $f(x, y) = ay^2 + 2bx^2y + cx^4 + R_5(x, y)$, where a is positive; in this case we have

$$\frac{\partial G_4}{\partial y} = 2(ay + bx^2),$$

* See Stolz, *Grundzüge*, vol. I, p. 234.

and this vanishes for $y=-\frac{b}{a}x^2$. We have

$$G_4(x, -x)=ax^2-2bx^3+cx^4, \quad G_4(x, x)=ax^2+2bx^3+cx^4,$$

and

$$G_4\left(x, -\frac{b}{a}x^2\right)=\frac{ac-b^2}{a}x^4.$$

It follows that $G_4(x, -x)$ or $G_4(x, x)$ is the value of $G_4(x, \overline{\phi}(x))$, and that $G_4\left(x, -\frac{b}{a}x^2\right)$ is that of $G_4(x, \underline{\phi}(x))$. If $ac-b^2$ be negative, the two expressions $G_4(x, \overline{\phi}(x))$, $G_4(x, \underline{\phi}(x))$ have opposite signs; therefore $f(0, 0)$ is not an extreme of $f(x, y)$. If $ac-b^2$ be positive, the two expressions are both positive, and the first condition of the general theorem is satisfied, since the indices of x in the leading terms are not greater than 4.

We find that $\frac{\partial G_4}{\partial x}=0$ has for roots $x=\pm\sqrt{-\frac{by}{c}}$, and $x=0$; we thus form the expressions

$$G_4(0, y)=ay^2, \quad G_4(\pm y, y)=ay^2+2by^3+cy^4.$$

It is unnecessary to consider the roots $x=\pm\sqrt{-\frac{by}{c}}$, because, for sufficiently small values of y, $|x|>|y|$, and thus these roots could not give the extremes for $|x|\leqq|y|$. Remembering that a and c are both positive, let $b\geqq 0$, then the value of $G_4(\overline{\psi}(y), y)$ is $ay^2+2by^3+cy^4$, and that of $G_4(\underline{\psi}(y), y)$ is ay^2; these values being both positive, we see that $G_4(0, 0)$ is a proper minimum of $G_4(x, y)$. The same conclusion may be made when $b\leqq 0$. Therefore, when $ac-b^2>0$, $a>0$, since the conditions of the theorem of § 370 are satisfied, $f(x, y)$ has a proper minimum at $(0, 0)$. If $ac-b^2>0$, $a<0$, there is a proper maximum. If $ac-b^2=0$, we have

$$f(x, y)=\frac{1}{a}(ay+bx^2)^2+R_5(x, y);$$

hence $G_4(x, y)$ has an improper extreme at $(0, 0)$, and no conclusion can be drawn as regards $f(x, y)$.

3.* Let $f(x, y)=y^2+x^2y+R_4(x, y)$. We find $\frac{\partial G_3}{\partial y}=2y+x^2$, and thence we have

$$G_3(x, -\tfrac{1}{2}x^2)=-\tfrac{1}{4}x^4;$$

also

$$G_3(x, x)=x^2+x^3, \quad G_3(x, -x)=x^2-x^3.$$

It is clear that, in this case, $G_3(x, \overline{\phi}(x))$, $G_3(x, \underline{\phi}(x))$ have opposite signs, provided x be sufficiently small, therefore $G(x, y)$ has no extreme at the point $(0, 0)$. Since

$$G_3(x, \underline{\phi}(x))=-\tfrac{1}{4}x^4,$$

it is not the case that $|G_3(x, \underline{\phi}(x))|\geqq c|x|^3$, for any value of c, in a neighbourhood of $x=0$; the theorem of § 370 is therefore not applicable. No information is obtained as to whether $f(x, y)$ has an extreme at $(0, 0)$, or not. It will in fact be shewn, in the next example, that $y^2+x^2y+x^4$ has a minimum at $(0, 0)$.

4. Let $f(x,y)=y^2+x^2y+x^4+R_5(x, y)$. We find $\frac{\partial G_4}{\partial y}=2y+x^2$, hence

$$\frac{\partial G_4}{\partial y}=0 \text{ gives } y=-\tfrac{1}{2}x^2;$$

hence

$$G_4(x, -\tfrac{1}{2}x^2)=\tfrac{3}{4}x^4+\dots;$$

also

$$G_4(x, x)=x^2+x^3+x^4, \quad G_4(x, -x)=x^2-x^3+x^4.$$

* Scheeffer, *loc. cit.*, p. 573.

In this case $G_4(x, \overline{\phi}(x))$, $G_4(x, \underline{\phi}(x))$ are both positive, and are greater than $c|x|^4$ for a fixed c. It can be shewn that the other condition is also satisfied. It follows that $f(x, y)$ has a minimum at $(0, 0)$.

5. Let $f(x, y)=x^2y^4-3x^4y^3+x^6y^2-3xy^7+y^8-10x^{10}y+5x^{12}+R_{13}(x, y)$.

In this case $\frac{\partial G_{12}}{\partial y}=0$ has the three roots

$$y=5x^4+\dots,\quad y=-\tfrac{1}{4}x^2-\tfrac{40}{7}x^4+\dots,\quad y=2x^2+\tfrac{5}{7}x^4+\dots.$$

On substituting these values in $G_{12}(x, y)$, and forming also $G_{12}(x, x)$, $G_{12}(x, -x)$, we find that $G_{12}(x, -\overline{\phi}(x))$ is $G(x, -x)$ or $G(x, x)$ according as x is positive or negative; and the expression commences with the term x^6. We find for $G_{12}(x, \underline{\phi}(x))$ an expression $-4x^{10}+\dots$. Since $G_{12}(x, \overline{\phi}(x))$, $G_{12}(x, \underline{\phi}(x))$ have opposite signs, it follows that $(0, 0)$ is not a point at which $G_{12}(x, y)$ has an extreme. Since the indices of the leading terms of $G_{12}(x, \overline{\phi}(x))$, $G_{12}(x, \underline{\phi}(x))$ are both less than 12, the condition of the theorem of § 370 is satisfied, and we can therefore infer that $f(x, y)$ has no extreme at $(0, 0)$.

FUNCTIONS REPRESENTABLE BY SERIES OF CONTINUOUS FUNCTIONS.

373. Before proceeding to consider the most general class of functions which are representable as the limits of sequences of continuous functions, and therefore by infinite series of which the terms are continuous functions, we shall first examine the case in which the function to be represented is itself continuous in a given interval.

The following general theorem is due to Weierstrass*:—

If a function $f(x)$ be continuous throughout a given interval (a, b), and if δ be an arbitrarily chosen positive number, a finite polynomial $P(x)$ can be found such that $|f(x)-P(x)|<\delta$, for every value of x in (a, b).

In order to prove the theorem, it is convenient first to consider certain special cases. Let a function y be defined, for the interval $(-\alpha, \alpha)$, by the specifications $y=mx$, for $0 \leqq x \leqq \alpha$, and $y=-mx$, for $-\alpha \leqq x \leqq 0$; thus y is the continuous function which is represented geometrically by portions of two straight lines which meet at the origin and are equally inclined to the x-axis. The function is represented in the whole interval $(-\alpha, \alpha)$ by

$$y=m\alpha\left[1+\left(\frac{x^2}{\alpha^2}-1\right)\right]^{\frac{1}{2}},$$

where the positive value of the radical is to be taken; the expression for y can be expanded by the Binomial Theorem in a series of powers of $\frac{x^2}{\alpha^2}-1$,

* *Sitzungsber.* of the Berlin Acad., 1885. The proof given here is substantially that due to Lebesgue; see *Bulletin des Sciences Math.*, ser. 2, vol. XXII (1), p. 278, 1898. Other proofs have been given by Picard, *Traité d'Analyse*, vol. I, p. 258; by Volterra, *Rend. del Circolo mat. di Palermo*, 1897, p. 83; by Mittag-Leffler, *Rend. del Circolo mat. di Palermo*, 1900; by Lerch, *Acta Math.*, vol. XXVII, p. 339. See also Borel's *Leçons sur les fonctions de variables réelles*, p. 50.

and this series converges uniformly in the whole interval. In this manner, by taking one, two, three, &c. terms of the series, we obtain a sequence of polynomials in x which converges uniformly to the value of the function; and thus this particular case of the general theorem has been established.

Next, let the function y be defined for the interval (a, b) as follows:—

Let $y = 0$, for $a \leqq x \leqq c$, and $y = m(x - c)$, for $c \leqq x \leqq b$, where c is a fixed number between a and b; this function is represented geometrically by the portion of the x-axis between the points a and c, and by the portion of the straight line $y = m(x - c)$ between the points c and b. The function may be represented by

$$y = \frac{m}{2}(x - c) + \left|\frac{m}{2}(x - c)\right|,$$

and since, as has been shewn in the last case,

$$\left|\frac{m}{2}(x - c)\right|$$

is representable as the limit of a sequence of polynomials, the same is true of the function now considered.

Next, let (a, b) be divided into a finite number of intervals

$$(a, x_1),\ (x_1, x_2),\ (x_2, x_3), \ldots (x_{n-1}, b),$$

and let ordinates to the x-axis be erected at the points

$$a,\ x_1,\ x_2, \ldots x_{n-1},\ b,$$

the extremities of these ordinates being denoted by

$$P,\ P_1,\ P_2, \ldots P_{n-1},\ Q.$$

Let the consecutive pairs of these points be joined by straight lines, an open polygon $PP_1P_2 \ldots Q$ being thus formed; it will be shewn that the continuous function $\phi(x)$ defined by the ordinates of this open polygon is such that a polynomial $P(x)$ can be found, such that

$$|\phi(x) - P(x)| < \eta,$$

for every value of x in (a, b). It is clear that the function $\phi(x)$ can be expressed as the sum of n functions

$$\phi_1(x),\ \phi_2(x), \ldots \phi_n(x),$$

which are such that $\phi_1(x)$ is linear in the whole interval (a, b); $\phi_2(x)$ vanishes in the interval (a, x_1), and is linear in the interval (x_1, b); $\phi_3(x)$ vanishes in the interval (a, x_2), and is linear in the interval (x_2, x_n); and generally $\phi_r(x)$ vanishes in the interval (a, x_{r-1}), and is linear in the interval (x_{r-1}, b). Since polynomials satisfying the prescribed conditions can be found for each of the functions

$$\phi_1(x),\ \phi_2(x), \ldots \phi_n(x),$$

the theorem is established for the function $\phi(x)$.

In the general case in which $f(x)$ is any function continuous in (a, b), it follows from the known theorem that $f(x)$ is necessarily uniformly continuous in (a, b), that, if ϵ be a prescribed positive number, the interval can be divided into parts $(a, x_1), (x_1, x_2), \ldots (x_{n-1}, b)$, such that the fluctuation of $f(x)$ in each of these parts is $< \epsilon$.

If $\phi(x)$ denotes the function considered above, which we take to be equal to $f(x)$ at each of the points $a, x_1, x_2, \ldots b$, and to be linear between each consecutive pair of these points, we see that

$$|f(x) - \phi(x)| < \epsilon$$

in the whole interval; and as it has been shewn that a polynomial $P(x)$ exists, such that

$$|\phi(x) - P(x)| < \eta,$$

it follows that

$$|f(x) - P(x)| < \epsilon + \eta;$$

hence, since ϵ, η are both arbitrarily chosen, Weierstrass' theorem has been established.

If $\delta_1, \delta_2, \ldots \delta_n, \ldots$ be a diminishing sequence of positive numbers which converges to the limit zero, then a sequence of polynomials

$$P_1(x),\ P_2(x), \ldots$$

can be found such that

$$|f(x) - P_1(x)| < \delta_1,\ |f(x) - P_2(x)| < \delta_2, \ldots |f(x) - P_n(x)| < \delta_n, \ldots$$

for all values of x in (a, b).

Since the sequence of polynomials $P_1(x), P_2(x), \ldots P_n(x), \ldots$ converges uniformly to $f(x)$ as their limit, $f(x)$ may be regarded as the sum-function of the uniformly convergent series

$$P_1(x) + [P_2(x) - P_1(x)] + \ldots + [P_n(x) - P_{n-1}(x)] + \ldots;$$

thus the following theorem has been established:—

If $f(x)$ be a function which is continuous throughout the interval (a, b), the function is expressible as the limiting sum of a uniformly convergent series, of which the terms are finite polynomials.

Weierstrass' theorem may be extended to the case of functions of any number of variables. The general result may be stated as follows:—

A function defined for any closed domain D, and continuous in that domain with respect to $(x_1, x_2, \ldots x_n)$, can be represented in that domain by a series of polynomials which converges uniformly and absolutely in D.

For a proof of this theorem, and for a discussion of the methods of Lagrange and Tchebicheff for the approximate representation of functions by series of polynomials, reference may be made to Borel's *Leçons sur les fonctions de variables réelles*, Chapter IV.

374. The question which arises as to the nature of the most general function that can be represented in a given interval as the sum of a series of continuous functions which converges at every point in the given interval, has been completely answered by Baire, whose result is contained in the following general theorem*:—

The necessary and sufficient condition that a function may, in a given interval, be representable as the sum of a series of continuous functions which converges at every point to the value of the function, is that the given function shall be at most point-wise discontinuous with respect to every perfect set of points in the given interval.

It will be in the first place assumed that the given function is limited in its domain.

To shew that the condition stated in the theorem is necessary, let

$$u_1(x) + u_2(x) + \dots + u_n(x) + \dots$$

be a series, such that the functions $u_n(x)$ are all continuous in a given interval (a, b), and such that the series converges for every point in (a, b). Instead of the function $s_n(x)$, we may consider the transformed sum-function $s(x, y)$, as defined in § 349. This function is defined for every point in the rectangle bounded by the four straight lines $x = a$, $x = b$, $y = 0$, $y = 1$, and is everywhere continuous with respect to y; it is also everywhere continuous with respect to x, except upon the axis $y = 0$, upon which its properties are to be investigated. The function $s(x, y)$ falls under the case investigated in § 244, where it is shewn that in every portion of a continuous curve

$$y = \phi(x),$$

there exist points at which $s(x, y)$ is continuous with respect to (x, y). In particular, on the axis $y = 0$, there must in every interval be points at which $s(x, y)$ is continuous with respect to (x, y), and *a fortiori* with respect to x; thus the function $s(x, 0)$ which represents the limiting sum of the series is at most point-wise discontinuous with respect to x. The same result holds if only such values of x are taken into account, as correspond to the points of any perfect set in the interval (a, b); hence it has been shewn to be a necessary property of the sum of a convergent series of continuous functions that it is at most point-wise discontinuous with respect to every perfect set.

* The theorem is fully developed in Baire's memoir, "Sur les fonctions de variables réelles," *Annali di Mat.* ser. III. A, vol. III, 1899; also in his treatise *Leçons sur les fonctions discontinues.* He has proved the sufficiency of the condition more succinctly in the *Bulletin de la Soc. Math. de France*, vol. XXVIII (1900); this is the proof given in the text. Another proof of the theorem, of a very instructive character, has been given by Lebesgue in Borel's *Leçons sur les fonctions de variables réelles*, pp. 149—155. See also Lebesgue's memoir "Sur les fonctions représentables analytiquement," *Liouville's Journal*, ser. 6, vol. I, 1905, where the theorem is proved, and also generalised.

In order to prove that the condition stated in the theorem is sufficient, let us suppose that a function $f(x)$ is defined for the domain (0, 1), and is at most point-wise discontinuous with respect to every perfect set of points contained in that domain; it must then be shewn that a sequence

$$f_1(x), f_2(x), \dots f_n(x) \dots$$

of functions can be determined which converges at every point of (0, 1) to the value of $f(x)$.

If p be a positive integer, the points

$$x = \frac{\alpha_\iota}{2^p},$$

where α_ι is a positive integer, will be called principal points of order p; the interval

$$\left(\frac{\alpha_\iota - 1}{2^p}, \frac{\alpha_\iota + 1}{2^p}\right)$$

of which the centre is $\frac{\alpha_p}{2^p}$, will be called a principal domain D_p of order p; the domains with centres 0, 1, of order p, will be taken to be the intervals

$$\left(0, \frac{1}{2^p}\right), \left(1 - \frac{1}{2^p}, \frac{1}{2^p}\right) \text{ respectively.}$$

A point P of (0, 1) belongs to two domains of order p; the centres of these domains will be called the principal points of order p associated with P.

The problem of constructing the required sequence of functions is reducible to the following problem:—

(A) To determine a number $\phi(D)$, corresponding to each principal domain D, which is such that, having given any point P and a positive arbitrarily chosen number ϵ, there exists an interval with centre P, such that, for every principal domain D entirely contained in the interval, and containing P, the number $\phi(D)$ differs from $f(P)$ by less than ϵ.

If the numbers $\phi(D)$ have been determined, the functions f_p can be constructed as follows:—At each principal point Q of order p, let f_p have for its value that of $\phi(D)$ corresponding to the domain of which Q is the centre. Then let f_p be made continuous, and such as to have at every point P a value intermediate between the greater and the lesser of its values at the principal points of order p, associated with P; for example f_p may be taken to be linear with respect to x, between each consecutive pair of principal points. Now take an interval with centre P satisfying the condition stated above; when p exceeds a certain value, the principal domains of order p, which contain P, are interior to this interval; hence the values of f_p at the principal points associated with P differ from $f(P)$ by less than ϵ; and the same is true of the function $f_p(P)$. Hence it follows that $f(P)$ is the limit of $f_p(P)$, when p is indefinitely increased.

Before we proceed to shew that the problem (A) can be solved for the case of any function f which is at most point-wise discontinuous relatively to every perfect set of points, it will be shewn that the problem is capable of a very simple solution in the case in which f is a semi-continuous function. Let us suppose that f is an upper semi-continuous function. In each domain D, let $\phi(D)$ be the maximum of f in that domain, then the conditions of (A) are satisfied; for an interval with P as centre can be found in which, at every point P', $f(P') < f(P) + \epsilon$. If D is contained in this interval, and contains P, the maximum of f in this interval, that is $\phi(D)$, lies between $f(P)$ and $f(P) + \epsilon$, hence the conditions of (A) are satisfied. It has thus been established that *every function f which is semi-continuous throughout* (0, 1) *is the limit of a sequence of continuous functions, and is therefore representable as the sum of an infinite series of continuous functions.*

To solve the problem (A) in the general case, let us take a descending sequence $\epsilon_1, \epsilon_2, \ldots \epsilon_n, \ldots$ of positive numbers which converge to zero; let D be a principal domain, then we have to define $\phi(D)$.

Let σ_1 be the greatest number in the sequence $\{\epsilon\}$, such that in D there are points at which the saltus of f is $\geqq \sigma_1$; let P_1 denote the set of points in D for which the saltus $\omega(f) \geqq \sigma_1$. If the perfect component P_1^{Ω} of P_1 exist, and if f be not continuous relatively to P_1^{Ω}, let σ_2 be the greatest number of the sequence $\{\epsilon\}$, such that there are points of D at which the saltus $\omega(f, P_1^{\Omega})$ of f relatively to P_1^{Ω}, is $\geqq \sigma_2$; let P_2 be the set of such points. In this manner we obtain closed sets

$$P_1, P_2, P_3, \ldots P_n, \ldots P_\omega, \ldots P_\alpha \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1),$$

all contained in D, and a corresponding set of numbers

$$\sigma_1, \sigma_2, \sigma_3, \ldots \sigma_n, \ldots \sigma_\omega, \ldots \sigma_\alpha \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2),$$

all belonging to the sequence $\{\epsilon\}$. If α be a non-limiting number, σ_α and P_α are derived from $\sigma_{\alpha-1}, P_{\alpha-1}$, just as σ_2 and P_2 are derived from σ_1, P_1. If α be a limiting number, P_α is the set common to all the sets of which the index is less than α, and σ_α is the inferior limit of all the σ's of which the index is less than α.

It is clear that, if $\beta < \alpha$, then $P_\beta \geqq P_\alpha$, and $\sigma_\beta \geqq \sigma_\alpha$.

It has been shewn in § 74, that, for such sets, there exists a number α of the first or of the second class, such that $P_\alpha = P_{\alpha+1}$. In the present case we must have $P_\alpha = 0$; for if P_α existed, we should have

$$P_\alpha = P_\alpha^{\Omega} = P_{\alpha+1},$$

and there would be a positive number $\sigma_{\alpha+1}$, such that, relatively to P_α^{Ω}, the saltus of f at every point is $\geqq \sigma_{\alpha+1}$; and thus f would be at every point discontinuous relatively to the perfect set P_α^{Ω}, which is contrary to the hypothesis made as regards the nature of the function f. It has thus been shewn that a number β exists, such that either P_β is enumerable, in which case $P_\beta^\Omega = P_{\beta+1} = 0$; or else such that P_β^Ω exists, but f is continuous relatively to it, in which case also $P_{\beta+1} = 0$.

If P_β be enumerable, there exists a number γ, such that P_β^γ consists of a finite number of points; we then take for $\phi(D)$ some number between the extreme values of f at the points of P_β^γ; it having been assumed that f is a limited function. If f is continuous relatively to P_β^Ω, we take for $\phi(D)$ any number between the extreme values of f in P_β^Ω.

The number $\phi(D)$ having now been defined for each domain, it must be shewn that the conditions of (A) are thus satisfied.

If P be a point of (0, 1), let τ_1 be the greatest of the numbers $\{\epsilon\}$ such that the saltus of f at P, $\omega_P(f) \geqq \tau_1$; and let Q_1 be the set of points in (0, 1) at which $\omega(f) \geqq \tau_1$. If Q_1^Ω exists and contains P, and if the function is not continuous at P relatively to Q_1^Ω, let τ_2 be the greatest number of the set $\{\epsilon\}$, such that $\omega_P(f, Q_1^\Omega) \geqq \tau_2$. Let Q_2 be the set of points in (0, 1) at which $\omega(f, Q_1^\Omega) \geqq \tau_2$; proceeding in this manner we obtain a set of closed sets

$$Q_1, Q_2, \ldots Q_n, \ldots Q_\omega, \ldots Q_\alpha \quad \text{......................(3)},$$

all of which contain P, and a set of numbers

$$\tau_1, \tau_2, \ldots \tau_n, \ldots \tau_\omega, \ldots \tau_\alpha \quad \text{......................(4)}.$$

The set Q_α and the number τ_α depend upon $Q_{\alpha-1}$, $\tau_{\alpha-1}$ just as Q_2, τ_2 depend upon Q_1, τ_1, in case α is not a limiting number. If α is a limiting number, Q_α consists of all points common to all the Q's with indices lower than α, and τ_α is the inferior limit of all the τ's with indices inferior to α. As before, Q_α must vanish after a certain index, and thus there is a number η, such that, either Q_η^Ω does not exist, or such that it exists and does not contain P, or else such that it contains P, but P is a point of continuity of f relatively to Q_η^Ω.

Each of the numbers of (4) belongs to the sequence $\{\epsilon\}$, except perhaps the last one τ_η; these numbers can be arranged in groups, the numbers of the same group being equal to the same number λ of the sequence $\{\epsilon\}$. The index of the first number in each group is necessarily a non-limiting number; for if α be a limiting number, as τ_α is the inferior limit of all the numbers $\tau_{\alpha'}$, such that $\alpha' < \alpha$, and as there exists only a finite number of values of the $\tau_{\alpha'}$, there are numbers α', such that $\tau_{\alpha'} = \tau_\alpha$. We can write

$$\left.\begin{array}{l} \tau_1 = \tau_2 = \ldots = \tau_{\alpha_1} = \lambda_1, \\ \tau_{\alpha_1+1} = \tau_{\alpha_1+2} = \ldots = \tau_{\alpha_2} = \lambda_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \tau_{\alpha_{k-1}+1} = \tau_{\alpha_{k-1}+2} = \ldots = \tau_{\alpha_k} = \lambda_k \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{array}\right\} \quad \text{......................(5)}.$$

The number of groups of the τ's may be finite, say k, in which case $\alpha_k = \eta$;

or else the number of groups is infinite, in which case $\lambda_1, \lambda_2, \dots \lambda_k, \dots$ have their limit zero.

Let $\lambda_1', \lambda_2', \dots \lambda_k', \dots$ be the numbers in the sequence $\{\epsilon\}$ which immediately precede the numbers $\lambda_1, \lambda_2, \dots \lambda_k, \dots$. Let R_1 be the set of points of (0, 1) at which $\omega(f) \geqq \lambda_1'$; let R_2 be the set of points of $Q_{\alpha_1}{}^\Omega$ for which

$$\omega(f, Q_{\alpha_1}{}^\Omega) \geqq \lambda_2';$$

and generally let R_k be the set of points of $Q^\Omega{}_{\alpha_{k-1}}$ for which

$$\omega(f, Q^\Omega{}_{\alpha_{k-1}}) \geqq \lambda_k'.$$

It may happen, if λ_1 be the first number of $\{\epsilon\}$, that no number λ_1' exists, and thus that R_1 may not exist. In accordance with the definition of the sets Q, the point P does not belong to any of the closed sets

$$R_1, R_2, \dots R_k, \dots;$$

hence, whatever h may be, there is an interval with centre P which contains no points of the sets

$$R_1, R_2, \dots R_h.$$

Let D be a principal domain containing P, and contained in the interval; if, in this domain, the sets $P_1, P_2, \dots P_\alpha$ are defined, all those of these sets of which the index is equal to or inferior to α_h coincide, in D, with the sets

$$Q_1, Q_2, \dots Q_\alpha,$$

having respectively the same indices; in fact, there is in D no point at which

$$\omega(f) \geqq \lambda_1',$$

whilst there is at least one, the point P, at which

$$\omega(f) \geqq \lambda_1 (= \tau_1),$$

therefore $\sigma_1 = \tau_1$; and consequently, in D, P_1 coincides with Q_1. It will be shewn that, in D, the sets P_α, Q_α coincide, if $\alpha \leqq \alpha_p$; it is sufficient to establish the proposition for a non-limiting number. There are two cases to be considered.

(1) If α be not the index of the first term of a group in (5); let us assume $\sigma_{\alpha-1} = \tau_{\alpha-1}$, and that $P_{\alpha-1}$, $Q_{\alpha-1}$ coincide in D. According to hypothesis $\tau_\alpha = \tau_{\alpha-1}$, so that at P the saltus relatively to

$$P^\Omega{}_{\alpha-1} = Q^\Omega{}_{\alpha-1}$$

is $\geqq \tau_\alpha$; therefore σ_α, which cannot exceed

$$\sigma_{\alpha-1} = \tau_{\alpha-1} = \tau_\alpha,$$

is identical with τ_α. Therefore P_α is, in D, identical with Q_α.

(2) If α be the index of the first term of one of the groups in (5), say $\alpha_\delta + 1$, $(\delta < h)$, we assume $P_{\alpha_\delta} = Q_{\alpha_\delta}$; in the set $Q^\Omega{}_{\alpha_\delta}$ the saltus at P is greater than or equal to $\lambda_{\delta+1}$. As D contains no point of $R_{\delta+1}$, the set of points where

$$\omega(f, Q^\Omega{}_{\alpha_\delta}) \geqq \lambda'_{\delta+1},$$

we have $$\sigma_{a_{s+1}} = \lambda_{s+1} = \tau_{a_{s+1}};$$

hence $$P_{a_{s+1}},\ Q_{a_{s+1}}$$

coincide in D. It has now been proved by induction that the two sets are coincident in D.

The final stage of the proof falls into two cases:—

(1) For the point P, the number of groups in (5) may be finite, say k; then $a_k = \eta$. The set Q_η exists, but $Q_{\eta+1}$ does not exist. There are then two sub-cases.

$(1)_a$ If $Q_\eta{}^\Omega$ does not exist, or exists but does not contain P; the point P belongs to a certain set $Q_\eta{}^r$, but not to $Q_\eta{}^{r+1}$; P is therefore an isolated point of $Q_\eta{}^r$. An interval can be determined, with P as its middle point, containing no point of $Q_\eta{}^r$ except P, and containing no points of $R_1, R_2, \dots R_k$. If D is a domain containing P, and contained in the interval, in this domain P_η and Q_η coincide; whence $P_\eta{}^r$, $Q_\eta{}^r$ consist of the one point P, and $P_\eta{}^{r+1} = 0$. In accordance with the definition of $\phi(D)$, we have $\phi(D) = f(P)$.

$(1)_b$ It may happen that P belongs to $Q_\eta{}^\Omega$, and that the function f is continuous at P relatively to $Q_\eta{}^\Omega$. An interval can be determined, with its centre at P, containing no points of $R_1, R_2, \dots R_k$, and such that the saltus of f relatively to the part of $Q_\eta{}^\Omega$ in the interval, is $< \epsilon$. If D is contained in the interval, and contains P, we have $P_\eta{}^\Omega$, $Q_\eta{}^\Omega$ coincident in D; the set $P_\beta{}^\gamma$ or $P_\beta{}^\Omega$ which occurs in the definition of $\phi(D)$, is certainly contained in $Q_\eta{}^\Omega$; hence $\phi(D)$ differs from $f(P)$ by less than ϵ.

(2) The groups in (5) may be infinite in number; if ϵ is fixed, k can be found such that $\lambda_k < \epsilon$. Relatively to $Q^\Omega{}_{a_k}$, the saltus at P is $< \epsilon$; thus an interval, with P as its middle point, can be determined, which contains no points of $R_1, R_2, \dots R_k$, and such that the fluctuation of f in the portion of $Q^\Omega{}_{a_k}$ which is contained in the interval is $< \epsilon$. As in $(1)_b$, the set $P_\beta{}^\gamma$ or $P_\beta{}^\Omega$ which serves to define $\phi(D)$, is contained in $Q^\Omega{}_{a_k}$; thus $\phi(D)$ differs from $f(P)$ by less than ϵ.

It has now been shewn that the conditions of the problem (A) are satisfied. Thus Baire's theorem has been completely established*, on the assumption that the given function is a limited one.

375. In order to extend the result to the case of a function which is not limited in its domain, let us suppose that the function $y = f(x)$ has points of infinite discontinuity; the case may also be included in which there are

* The above proof, as given by Baire, is applicable to the case of a function of any number of variables. The only modification required for this extension is an obvious generalisation in the definition of the principal points and principal domains.

values of x for which y has the improper value ∞, or the improper value $-\infty$. Let a new function $z = \phi(x)$ be defined by means of the relations

$$z = \frac{y}{1+y}, \text{ for } 0 \leqq y \leqq \infty,$$

and

$$z = \frac{y}{1-y}, \text{ for } -\infty \leqq y \leqq 0.$$

It is then clear that, corresponding to the unlimited function $y = f(x)$, we have a limited function $z = \phi(x)$, in which the dependent variable z has 1 and -1 for its upper and lower limits in the whole domain of x. It is easily seen that two values of y correspond to values of z in which the relation of order is conserved, and conversely; and further, that, to a convergent sequence of values of y, there corresponds a convergent sequence of values of z, and the converse. A point of continuity of $f(x)$ is also a point of continuity of $\phi(x)$, and the converse is true; and this also holds of the points of continuity with respect to any perfect set of points in the domain of x. It follows that, if one of the two functions $f(x)$, $\phi(x)$ is point-wise discontinuous with respect to every perfect set, then the other function has the same property. If $f(x)$ is the limit of a sequence $f_1(x)$, $f_2(x)$, ... of continuous functions, then the functions $\phi_1(x)$, $\phi_2(x)$, ... obtained by applying the above transformation are also continuous, and they converge to $\phi(x)$. Therefore, if $f(x)$ is the limit of a sequence of continuous functions, $\phi(x)$ has the same property. If $\phi(x)$ is assumed to be the limit of a sequence of continuous functions, we can assume these functions $\phi_n(x)$ to have $+1$, -1 for upper and lower limits, as these are the upper and lower limits of $\phi(x)$. Let $e_1, e_2, e_3, \ldots$ be a convergent sequence of increasing proper fractions, of which 1 is the limit.

The functions $e_1\phi_1(x)$, $e_2\phi_2(x)$, ... are continuous, and converge to the limit $\phi(x)$; moreover $e_n\phi_n(x)$ has $+e_n$, $-e_n$ for its upper and lower limits, and these numbers lie within the interval $(-1, +1)$. The function $f_n(x)$ which transforms into $e_n\phi_n(x)$ is then continuous and limited; and since the limit of the sequence $\{e_n\phi_n(x)\}$ is $\phi(x)$, it follows that the limit of the sequence $\{f_n(x)\}$ is $f(x)$. The condition that $\phi(x)$ is the limit of a sequence of continuous functions is that it should be at most point-wise discontinuous with respect to every perfect set in the domain of x; it follows that, when this condition is satisfied, $f(x)$ is also at most point-wise discontinuous with respect to every perfect set. The general theorem has now been completely established.

376. It may be shewn that any function $f(x)$ which satisfies the condition contained in Baire's general theorem can be expressed as the limit of a sequence of finite polynomials. Since, under the condition stated, $f(x)$ is the limit of a sequence $\{f_n(x)\}$ of continuous functions, if $\epsilon_1, \epsilon_2, \ldots$ be a sequence of decreasing positive numbers converging to zero, we can, in virtue of

Weierstrass' theorem (§ 373), determine a polynomial $P_n(x)$ such that $|P_n(x) - f_n(x)| < \epsilon_n$, in the whole domain of x. It follows that the sequence $P_1(x)$, $P_2(x)$, ... of polynomials so determined, converges to the limit $f(x)$. We have, in particular, the following theorem:—

The necessary and sufficient condition that a function $f(x)$ can be represented by a convergent series of polynomials, each of finite degree, is that $f(x)$ should be at most point-wise discontinuous with respect to every perfect set of points in the domain of x.

377. A theorem relating to those continuous functions which in any interval of the independent variable are either differentiable, or at least possess everywhere a definite derivative on one and the same side, can be deduced from Baire's general theorem. The derivative may at any point be infinite with a definite sign.

If $\frac{f(x+h) - f(x)}{h}$, for positive values of h have a definite limit, finite or infinite, for each value of x, as h has a sequence of positive values converging to zero, it follows from Baire's theorem, that the derivative so defined is at most point-wise discontinuous with respect to every perfect set in the interval. Thus the following theorem has been established:—

If $f(x)$ be continuous in any interval, and possess at every point a differential coefficient, or else a definite derivative on one and the same side, then the differential coefficient, or the derivative, is either continuous in the interval, or else is point-wise discontinuous with respect to every perfect set of points contained in the interval.

BAIRE'S CLASSIFICATION OF FUNCTIONS.

378. A classification of functions of a real variable has been suggested by Baire*, based upon the properties of the functions in relation to their representation as limits of sequences of functions. Continuous functions form the class 0; and functions which are not continuous but are the limits of sequences of continuous functions belong to the class 1. Functions of class 2 are those which can be represented as the limits of sequences of functions of class 1, and do not themselves belong to either of the classes 0 or 1. In general, a function is of class n, if it can be represented as the limit of a sequence of functions of class $n-1$, provided it does not belong to any of the classes 0, 1, 2, ... $n-1$.

It can be shewn by means of an example, that functions of class 2 exist. Consider the function $f(x)$ which has the value 1, for all rational values of x, and the value 0, for every irrational value of x. This function does not belong to either of the classes 0 and 1, for it is totally discontinuous;

* *loc. cit.* (above, p. 525). See also *Comptes Rendus*, December 4 and 11, 1899.

but it is easily seen to be the limit of a sequence of functions all of which are of class 1. For let us define $f_n(x)$ to be zero at every point except those points at which the value of x is rational and has for its denominator an integer not exceeding n; this function $f_n(x)$ has only a finite number of discontinuities in any given interval, and therefore belongs to class 1. The function $f(x)$ is the limit of the sequence $\{f_n(x)\}$, and is therefore of class 2. This function is capable of the analytical representation

$$f(x) = \lim_{m=\infty} \lim_{n=\infty} (\cos m!\, \pi x)^{2n}.$$

A function belonging to class 2 can be represented by a double series $\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} P_{m,n}(x)$, where $P_{m,n}$ denotes a finite polynomial. This double series cannot be reduced to a single one, the terms of which are continuous, for the function would then not be of class 2. In general, a function of class p must be representable by a p-fold series of finite polynomials.

A function which is the limit of a sequence of functions belonging to the classes 1, 2, 3, ... n, ..., and which does not itself belong to any of these classes, is said to be of class ω, where ω denotes the first transfinite ordinal number. A function which is the limit of a sequence of functions of class ω, and which is not itself of class ω, or of any class inferior to ω, is said to be of class $\omega + 1$. Proceeding in this manner, we may attach a meaning to the statement that a function is of class β, where β denotes any prescribed ordinal number of the second class. In case β be a limiting number, a function is said to be of class β when it is the limit of a sequence of functions each of which is of some class ordinally preceding β, provided the function be not itself of any of the classes preceding β. If β be a non-limiting number, a function which is the limit of a sequence of functions of class $\beta - 1$ is said to be of class β, provided it be not of any class inferior to β. A function belonging to any class is necessarily a summable function.

The question has been discussed by Baire* and by Borel†, whether it be possible effectively to define functions of all classes, the numbers of the classes being finite or transfinite. Those functions which belong to the classes of which the numbers are less than some prescribed number form an aggregate of cardinal number c, which is less than the cardinal number f of all functions. This would indicate that functions cannot be exhaustively classified under numbers less than some fixed number, either finite, or transfinite of the second class, but this does not settle the question whether it be possible to define a particular function which does not belong to one of the classes in question. The question also arises whether it be possible to define functions which do not belong to any of Baire's classes.

* *Annali di Mat.* ser. 3 A, vol. III, p. 71.

† *Leçons sur les fonctions de variables réelles*, p. 156.

These questions have been fully discussed by Lebesgue in his memoir* "Sur les fonctions représentables analytiquement." It is there pointed out that all functions that are representable analytically belong to Baire's classes. A function is said to be representable analytically when it is constructible by effecting, according to a determinate norm, a finite, or enumerably infinite, number of additions, multiplications, and passages to the limit, upon variables and constants. The other operations of analysis are reducible to those here enumerated. Lebesgue has advanced proofs that functions of all Baire's classes exist, in the sense that it is possible effectively to define a function of any prescribed class. He has also shewn that functions can be defined which do not belong to any of Baire's classes, and which are therefore incapable of analytical representation. In some of Borel's reasoning, correspondences are assumed to exist which are defined only by means of an infinity of separate acts of choice. The objection to such reasoning is recognised by Lebesgue†, who has endeavoured to avoid this difficulty. In view of the grave difficulties connected with the necessary nature of an adequate definition of a function, and which form the subject of controversy at present, it would perhaps be premature to assume that the questions here referred to have been finally settled.

THE INTEGRATION OF SERIES.

379. Let $u_1(x), u_2(x), \ldots u_n(x), \ldots$ be limited integrable functions defined for the interval (a, b), and such that, at each point of the domain, the series $u_1(x) + u_2(x) + \ldots + u_n(x) + \ldots$ converges to the sum $s(x)$; the conditions will be determined that the function $s(x)$ has a proper integral in the domain (a, b). The term "integrable" is here used in the sense employed by Riemann.

Since the functions $s_1(x), s_2(x), \ldots s_n(x), \ldots$ are limited functions, $|s_1(x)|$, $|s_2(x)|, \ldots |s_n(x)|, \ldots$ have upper limits $u_1, u_2, \ldots u_n, \ldots$ in the domain (a, b). If $u_1, u_2, \ldots u_n, \ldots$ have a finite upper limit U, it can be shewn that $|s(x)|$ has a finite upper limit in (a, b); for if the upper limit of $|s(x)|$ were indefinitely great, a value of x would exist such that $|s(x)| = U + \alpha$, where α is some positive number; now n can be taken so great that $|s(x) - s_n(x)| < \epsilon$, where ϵ is arbitrarily small, hence $s(x) < |s_n(x)| + \epsilon < U + \epsilon$, and since ϵ can be chosen to be $< \alpha$, it is impossible that $|s(x)| = U + \alpha$; and therefore it is impossible that the upper limit of $|s(x)|$ be not finite.

The condition just stated, that $|s(x)|$ may have a finite upper limit, is a sufficient one but not a necessary one; in fact we know that at the point $(x, 0)$, the function $s(x, y)$ may have an infinite discontinuity, whilst $s(x)$ has only a finite discontinuity, or is continuous, at the point x. In what follows,

* *Liouville's Journal*, ser. 6, vol. I, 1905.

† Also by Borel himself; see *Bulletin de la soc. math. de France*, vol. XXXIII, 1905, pp. 272, 273.

it will be assumed that $|s(x)|$ has a finite upper limit in (a, b), so that in case $s(x)$ is integrable, the integral is a proper one.

Let E be a set of points in (a, b) of measure zero. Let ϵ be an arbitrarily chosen positive number, and n_1 an arbitrarily chosen positive integer. Let us suppose that, for each point x_1 of (a, b) which does not belong to a certain component E_ϵ of E, an integer $n (> n_1)$ can be determined, and also a neighbourhood $(x_1 - \delta, x_1 + \delta')$, such that the condition $|R_n(x)| < \epsilon$ is satisfied for every point x in that neighbourhood and lying in (a, b). Then, provided this condition is satisfied for every value of ϵ, and E is such that each point of it belongs to E_ϵ, for some sufficiently small value of ϵ, the convergence of the sequence $s_1(x), s_2(x), \ldots$ to $s(x)$ is said to be *regular in* (a, b) *except for the set* E *of zero measure.*

It will be observed that, for a fixed ϵ, the integer $n (> n_1)$ depends in general upon the particular point x_1 which does not belong to E_ϵ. Moreover, since n_1 is arbitrary, there exists for a particular point x_1, an infinite number of values of n; the neighbourhood $(x_1 - \delta, x_1 + \delta')$ depending however in general upon the value of n chosen.

In the particular case in which every $u_n(x)$ is positive or zero, for every value of x and n, so that the sequence $s_1(x), s_2(x), \ldots$ is a non-diminishing sequence, when the condition $|R_n(x)| < \epsilon$ is satisfied for a particular value of n, it is also satisfied for every greater value. In the general case this does not hold; the condition is satisfied for an infinite number of greater values of n, but not necessarily for every such value.

It is easily seen that the set E_ϵ must, for each value of ϵ, be a non-dense closed set, although the set E is not necessarily non-dense, and may be everywhere-dense in (a, b). For, if ξ be a limiting point of the set E_ϵ, then every neighbourhood of ξ contains points of E_ϵ, and it is impossible that the condition $|R_n(x)| < \epsilon$ can be satisfied for every point of such a neighbourhood. Therefore ξ must itself belong to E_ϵ, which must consequently be a closed set; and since it has the measure zero, it cannot contain all the points of any interval (α, β), and is thus non-dense in (a, b).

380. The following theorem will now be established:—

The necessary and sufficient condition that the limited function $s(x)$ *may be integrable in* (a, b), *in accordance with Riemann's definition, is that the sequence of integrable functions* $s_1(x), s_2(x), \ldots$ *shall converge to* $s(x)$ *regularly, except for a set of points* E *of zero measure, and of the first category.*

To prove that the condition stated is necessary, let it be assumed that $s(x)$ is integrable in (a, b). The number ϵ, and the integer n_1 being fixed, let it be assumed that, if possible, the set E_ϵ of points of (a, b), for each of which it is impossible to fix a value of $n (> n_1)$, and a neighbourhood such that $|R_n(x)| < \epsilon$, for all points of that neighbourhood, has a measure greater than zero. Since $s(x)$, and $s_1(x)$, $s_2(x)$, ... are all integrable, the set of points at which one of these functions is discontinuous has the measure zero, and it follows that the set of all points at which one at least of these functions is discontinuous has the measure zero. Remove from E_ϵ every point at which one or more of the functions $s(x)$, $s_1(x)$, $s_2(x)$, ... is discontinuous, we then have left a set F_ϵ, of the same measure as E_ϵ, which is by hypothesis greater than zero. At every point of F_ϵ, the functions $s_n(x)$, and the function $s(x)$ are all continuous.

If ξ be a point of F_ϵ, the number $n\,(> n_1)$ can be so chosen that

$$|s(\xi) - s_n(\xi)| < \tfrac{1}{3}\epsilon;$$

also δ can be so chosen that, for every x in the interval $(\xi - \delta, \xi + \delta)$ the inequalities

$$|s(\xi) - s(x)| < \tfrac{1}{3}\epsilon, \quad |s_n(\xi) - s_n(x)| < \tfrac{1}{3}\epsilon$$

are both satisfied. This follows from the fact that $s(x)$, $s_n(x)$ are continuous at ξ. From these inequalities we deduce that the inequality

$$|s(x) - s_n(x)| < \epsilon$$

is satisfied at all points x in the interval $(\xi - \delta, \xi + \delta)$. But this is contrary to the hypothesis that ξ belongs to the set E_ϵ, for the points of which no neighbourhoods in which the last condition is satisfied can be determined. It therefore follows that it is impossible that the set E_ϵ can have its measure greater than zero.

The set E_ϵ having been now shewn to have the measure zero, we may consider a descending sequence $\epsilon_1, \epsilon_2, \epsilon_3, \ldots$ of values of ϵ converging to zero. The sets $E_{\epsilon_1}, E_{\epsilon_2}, E_{\epsilon_3}, \ldots$ have their measures zero, and they determine a set E of the first category, consisting of all points which belong to any of these sets. It follows from the theorem of § 82, that the set E has zero measure; and it has thus been established that, if $s(x)$ be integrable, then the convergence is regular except for the points of this set E.

To shew that the condition stated in the theorem is sufficient, let ϵ and n_1 be fixed, then the set E_ϵ is a non-dense set of zero content. The points of E_ϵ can therefore all be enclosed in the interiors of intervals of a finite set, the sum of whose lengths is an arbitrarily small number η. The remainder of (a, b) consists of a finite set of intervals; and for each point x_1 of any one of these intervals, a neighbourhood $(x_1 - \delta, x_1 + \delta')$ can be determined, and also a number $n\,(> n_1)$, not necessarily the same for all points x_1, such that the condition $|R_n(x)| < \epsilon$ is satisfied for all points of $(x_1 - \delta, x_1 + \delta')$ which are in (a, b). To the set of all such intervals we may apply the Heine-Borel theorem; and consequently a finite set of intervals can be determined, such that each point of (a, b), not in the interior of the excluded intervals, is in the interior of at least one of the intervals; and in each one of this finite set of intervals the condition $|R_n(x)| < \epsilon$ is everywhere satisfied for some one value of n, greater than n_1. When the set of intervals of which the sum is η is excluded from (a, b), the remainder may be divided into a finite number of parts such that, in each part, the condition $|R_n(x)| < \epsilon$ is satisfied for a value of n belonging to a finite set $n_1 + p_1, n_1 + p_2, \ldots n_1 + p_r$ of numbers all $> n_1$. To shew that $s(x)$ is integrable in (a, b), we now apply Riemann's test of integrability. Divide (a, b) into a number of parts $h_1, h_2, \ldots h_s$, so chosen that all the end-points of the excluded intervals, and also all the end-points of those finite parts for each of which $|R_n(x)| < \epsilon$ for a single value of n, are end-points of the parts $h_1, h_2, \ldots h_n$. For an interval h in the excluded set, the product of h into the fluctuation of $s(x)$ is less than $(M - m)h$, where M and m are the upper and lower limits of $s(x)$ in (a, b). For an interval h, for the whole of which $|R_{n_1+p}(x)| < \epsilon$, we see that the fluctuation of $s(x)$ cannot exceed that of $s_{n_1+p}(x)$ by more than 2ϵ. It follows that the sum of the products of each h into the corresponding fluctuation of $s(x)$ is not greater than

$$(M - m)\eta + \sum_p \Sigma h \{2\epsilon + \text{fluctuation of } s_{n+p}(x)\}$$

where, in the double summation, the first summation refers to all those of the h's which are in an interval for which p has one and the same value, and the second summation refers to the values $p_1, p_2, \ldots p_r$. Since $s_{n+p}(x)$ is integrable through the interval to which it belongs, and for which p has a fixed value, we see that when the number s is sufficiently increased, and the greatest of the h's is sufficiently small, $\sum_p \Sigma h \times$ fluctuation of $s_{n+p}(x)$ becomes arbitrarily small. Since η and ϵ are arbitrarily small, it follows that Riemann's test of integrability of $s(x)$ is satisfied.

The general theorem having now been completely established, it is seen from the foregoing proof that it may be stated in the following form:—

If $u_1(x) + u_2(x) + \ldots$ converges to a definite value $s(x)$ at every point in (a, b), and if all the functions $u_1(x), u_2(x), \ldots$ have proper integrals in (a, b), then the necessary and sufficient conditions that $s(x)$ may have a proper integral are (1) *that the upper limit of $|s(x)|$ in (a, b) be finite, and* (2) *that, corresponding to two arbitrarily small positive numbers η, ϵ, and to any positive integer n_1, a finite number of intervals whose sum is less than η can be excluded from (a, b), so that, in the remainder of (a, b), $|R_{n_1+p}(x)| < \epsilon$, for every x, where p has one of a finite number of values which depend on x, but are such that the same p is applicable to all points x in a certain continuous interval.*

The condition (2) contained in this theorem was obtained* first by Arzelà, and is expressed by him in the form, that there must be a certain mode of

* "Sulle serie di funzioni," Part II, *Mem. della R. Accad. d. Sci. di Bologna*, ser. 5, vol. VIII, 1900. A proof different from that in the text was given by Hobson, see *Proc. Lond. Math. Soc.* ser. 2, vol. I, where it is shewn that Arzelà's proof of his theorem is invalid.

convergence of the series called ***uniform convergence by segments in general*** (convergenza uniforme a tratti in generale). This mode of convergence differs from that of uniform convergence by segments, considered in § 354, in that a finite number of intervals of arbitrarily small sum must be excluded from the domain in order that the condition may be satisfied.

381. The theorem obtained above contains the necessary and sufficient conditions that the limit of a sequence of integrable functions is itself integrable, the conception of integration being that of Riemann. The corresponding theorem, when the conception of an integral in the extended sense introduced by Lebesgue is employed, is of a simpler form:—

If $u_1(x), u_2(x), \ldots u_n(x), \ldots$ *be limited functions, defined for* (a, b), *which are integrable in accordance with Lebesgue's definition, and if the series* $u_1(x) + u_2(x) + \ldots + u_n(x) + \ldots$ *converge to the function* $s(x)$, *limited in the whole interval* (a, b), *then the function* $s(x)$ *is also integrable in accordance with Lebesgue's definition.*

Let A, B be any two fixed numbers, such that $A < B$, and let ϵ be an arbitrarily chosen positive number. Let us consider the sets of points $G_n(\epsilon)$, $G_{n+1}(\epsilon), \ldots G_{n+m}(\epsilon), \ldots$, where n is a fixed integer, and $G_{n+m}(\epsilon)$ denotes the set of points x, such that $A - \epsilon < s_{n+m}(x) < B + \epsilon$; these sets are all measurable, since all the functions $s_{n+m}(x)$ are integrable in accordance with Lebesgue's definition of integration. The set of points x for which $A \leqq s(x) \leqq B$, is such that any point x of the set belongs to all the sets $G_n(\epsilon), G_{n+1}(\epsilon), \ldots G_{n+m}(\epsilon), \ldots$, from and after some fixed one of the sets. Denoting by $G(\epsilon)$ the set of all points each of which belongs to all the sets $G_n(\epsilon), G_{n+1}(\epsilon), \ldots$, from and after some fixed one of the sets, we see that the set of points x such that $A \leqq s(x) \leqq B$, is a component of the set $G(\epsilon)$, which set $G(\epsilon)$ is measurable (§ 82). Now take a descending sequence $\epsilon_1, \epsilon_2, \ldots \epsilon_s, \ldots$ of values of ϵ, which converges to zero; the set of points such that $A \leqq s(x) \leqq B$, is the set which is common to all the measurable sets $G(\epsilon_1), G(\epsilon_2), \ldots G(\epsilon_s) \ldots$; and this set is therefore itself measurable. Since the set of points for which $A \leqq s(x) \leqq B$ is measurable whatever values A and B may have, and $s(x)$ is a limited function, it follows that $s(x)$ satisfies Lebesgue's condition of integrability.

The theorem may be also stated as follows:—

If the function $s(x)$ *be the limit of a sequence* $\{s_n(x)\}$ *of limited summable functions, defined for an interval* (a, b), *then* $s(x)$ *is also a summable function, and therefore has a Lebesgue integral in* (a, b), *in case it be limited.*

382. When the sum-function $s(x)$ of the series $\Sigma u(x)$ is integrable in (a, b), and therefore in any interval (a, x), where $x \leqq b$, the question arises whether the sum $\sum_{n=1}^{\infty} \int_a^x u_n(x)\,dx$ is finite and continuous, and if so whether it is equal to $\int_a^x s(x)\,dx$, which is necessarily a continuous function. Sufficient conditions will be obtained that such a term by term integration of the series $\Sigma u(x)$ is valid.

The following theorem will be first established:—

If the series $\Sigma u(x)$ converge to $s(x)$ for every value of x in (a, b), and if ϵ be a fixed positive number, the set of points G_n for which $|R_n(x)| > \epsilon$, is such that its interior measure has the limit zero, when n is indefinitely increased.

To prove this theorem we observe that, if the limit of the interior measure of G_n is not zero, there must be an infinite number of values of n for which the interior measure of G_n is greater than some number α. It follows from the theorem of § 93, that a set of points of interior measure $\geqq \alpha$ exists, each of which belongs to an infinite number of the sets G_n. This is inconsistent with the condition that the series $\Sigma u(x)$ converges for each value of x; for any fixed point x, there can only be a finite number of values of n such that $|R_n(x)| > \epsilon$. It follows that the interior measure of G_n must have the limit zero, when n is indefinitely increased.

383. Let us now assume that $s(x)$ has a proper integral, and further that $|R_n(x)|$ is, for every value of n and of x, less than some fixed number C; then, $R_n(x)$ being integrable, $|R_n(x)|$ is also integrable, and therefore the set of points for which $|R_n(x)| > \epsilon$ is measurable. It follows from the theorem proved above, that the measurable set of points for which $|R_n(x)| > \epsilon$, is such that its measure has the limit zero, when n is indefinitely increased. We see therefore that

$$\left|\int_a^x R_n(x)\,dx\right| < (x-a)\,\epsilon + \eta C,$$

where n is chosen so great that the measure of the set of points, at each of which $|R_n(x)| > \epsilon$, is less than the arbitrarily chosen number η. Since ϵ, η are arbitrarily small, it follows that

$$\lim_{n=\infty}\int_a^x R_n(x)\,dx = 0,$$

and thus that

$$\int_a^x s(x)\,dx = \lim_{n=\infty}\int_a^x s_n(x)\,dx.$$

The following theorem has now been established:—

If $u_1(x) + u_2(x) + \ldots + u_n(x) + \ldots$ be a series of limited integrable functions which converges at every point in the interval (a, b) to the integrable function $s(x)$, then if $|R_n(x)| \equiv |s(x) - s_n(x)|$ is, for every value of x

and of n, less than some fixed number C, the series may be integrated term by term in any interval (a, x), the sum of the integrals converging to

$$\int_a^x s(x)\,dx.$$

The condition stated in this theorem may be replaced by the condition that $|s_n(x)|$ be less than some fixed number, for all values of n and x.

If the transformed remainder function $|R(x, y)|$ have no upper limit in its domain, there must be at least one point such that the saltus of the function is indefinitely great; it is clear that such a point must be in the x-axis, and is a point at which the measure of non-uniform convergence is indefinitely great. Conversely, if there is no such point, the upper limit of $|R(x, y)|$, or of $|R_n(x)|$, is finite. The foregoing theorem may thus be stated as follows*:—

A sufficient condition for the term by term integrability of a series of limited integrable functions which converges in a given interval to an integrable function s(x), is that there be no points at which the measure of non-uniform convergence of the series is indefinitely great.

A special case of the general result which has been obtained, is that in which the series converges uniformly. This condition is sufficient to ensure both that the sum-function $s(x)$ is integrable, and that its integral through any interval (a, x) is the sum of the corresponding integrals of the functions $u(x)$; thus we have the theorem:—

If a series $\Sigma u(x)$ of integrable functions converge uniformly in the interval (a, b) to the sum s(x), then the sum $\Sigma \int_a^x u(x)\,dx$ converges to the value $\int_a^x s(x)\,dx$, where $a \leqq x \leqq b$.

If it be assumed that the convergence of $\Sigma u(x)$ is simply uniform only, this is sufficient to ensure that $s(x)$ is integrable, but it is then not necessarily true that $\Sigma \int_a^x u_n(x)\,dx$ is a convergent series. It can however be shewn that whenever this series is convergent, it converges to the value

$$\int_a^x s(x)\,dx.$$

In fact we know that, by bracketing the terms of the simply uniformly convergent series $\Sigma u_n(x)$ in a suitable manner, the series is thereby converted

* This theorem was obtained first by Osgood, for the case in which $s(x)$ and all the $u(x)$ are continuous; see *Amer. Journal of Math.*, vol. XIX, 1897. The case in which $s(x)$ is not necessarily continuous was obtained by Hobson, *Proc. Lond. Math. Soc.*, vol. XXXIV, p. 245, and the general case was investigated by W. H. Young, *Proc. Lond. Math. Soc.*, ser. 2, vol. I, and also by Arzelà, *loc. cit.*

into a uniformly convergent series $\Sigma v_m(x)$, and the above theorem is then applicable to the new series, and thus $\sum_1^m \int_a^x v_m(x)\,dx$ converges to the value $\int_a^x s(x)\,dx$. It is clear that whenever $\sum_1^n \int_a^x u_n(x)\,dx$ converges, it must converge to the same value as does the series $\sum_1^m \int_a^x v_m(x)\,dx$. We thus obtain the following theorem:—

If a series $\Sigma u(x)$ of integrable functions converge simply-uniformly in the interval (a, b) to the sum $s(x)$, then (1) if the series $\Sigma \int_a^x u(x)\,dx$ be convergent, it converges to the value $\int_a^x s(x)\,dx$, and (2) if the series be not convergent, it may by suitably bracketing the terms, and amalgamating the terms in each bracket, be converted into a series which converges to the value $\int_a^x s(x)\,dx$.*

384. It has been shewn in § 383, that if $|R_n(x)|$ is $< C$, for every value of n and of x in the interval (a, b), then

$$\left|\int_a^x R_n(x)\,dx\right| < (x-a)\,\epsilon + \eta C,$$

where n is so great that the measure of the set of points in (a, x) at which $|R_n(x)| > \epsilon$, is $< \eta$, where ϵ, η denote arbitrarily chosen numbers. Since the set G_n of points of (a, b) at which $|R_n(x)| > \epsilon$, has the limit zero when n is indefinitely increased, we may fix a number n_1 such that, for every value of n that is $\geqq n_1$, the measure of G_n is less than η. We have then

$$\left|\int_a^x R_n(x)\,dx\right| < (x-a)\,\epsilon + \eta C < (b-a)\,\epsilon + \eta C,$$

provided $n \geqq n_1$, for every value of x in the interval (a, b). Since $(b-a)\,\epsilon + \eta C$ is arbitrarily small, it follows that $\int_a^x R_n(x)\,dx$ converges uniformly to zero for all values of x in the interval (a, b), as n is indefinitely increased. We have therefore the theorem†:—

When there are no points at which the measure of non-uniform convergence of the series $\Sigma u(x)$, of integrable functions, which converges to the integrable function $s(x)$ in (a, b), is indefinitely great, then the convergence of the series $\Sigma \int_a^x u(x)\,dx$ to the value $\int_a^x s(x)\,dx$ is uniform in the interval (a, b).

* The first part of this theorem was given by Bendixson, for the case in which the functions $u(x)$ are continuous; see *Stockholm Öfv.* vol. LIV, p. 609.

† This theorem was stated, and proved otherwise by W. H. Young; see *Comptes Rendus*, vol. CXXXVI, p. 1632.

The proofs of the two theorems in § 383 are still applicable when the functions are integrable only in the extended sense employed by Lebesgue. We have therefore the following theorem which includes the former ones as special cases:—

If $u_1(x), u_2(x), \ldots u_n(x), \ldots$ *be a sequence of limited functions integrable in* (a, b) *in the extended sense of the term, and if the series* $\Sigma u(x)$ *converges to* $s(x)$, *then in case* $|R_n(x)|$ *has a finite upper limit for all values of* n *and* x, *the series* $\Sigma \int_a^x u_n(x)\,dx$, *converges uniformly for all values of* x *in* (a, b) *to the sum* $\int_a^x s(x)\,dx$, *the integrals being taken to be Lebesgue integrals.*

This theorem may also be stated as follows:—

If $s(x)$ *be the limit of a sequence* $\{s_n(x)\}$ *of summable functions defined for the interval* (a, b), *and if* $|s_n(x)|$ *have a finite upper limit for all values of* n *and* x, *then* $\int_a^x s_n(x)\,dx$ *converges uniformly, for all values of* x *in* (a, b), *to* $\int_a^x s(x)\,dx$, *which has been shewn in* § 381 *to exist.*

385. We proceed to consider the case in which the condition that $|R_n(x)|$ has a finite upper limit for all values of x in (a, b), and all values of n, is not satisfied. In this case there is a set G of points at which the measure of non-uniform convergence of the series is indefinitely great; it has been shewn in § 350 that the set G is closed. If we assume that the series $\sum_1^\infty \int_a^x u_n(x)\,dx$ is everywhere convergent, and has $U(x)$ for its sum, and that the integral $\int_a^x s(x)\,dx$ everywhere converges to $S(x)$, it may happen that $U(x)$ is discontinuous, and is consequently not everywhere equal to the essentially continuous function $S(x)$. It may however happen that $U(x)$ is continuous, and yet is not equal to $S(x)$. It will however be shewn that, in case $U(x)$ is continuous, it is a sufficient condition for the equality of $S(x)$ and $U(x)$, that the set G should be enumerable.

Let x be a point which does not belong to G; then a neighbourhood $(x-\epsilon_1, x+\epsilon_2)$ of x can be found, such that, for all points in this neighbourhood, and for all values of n, $|R_n(x)|$ has a finite upper limit. Denoting the sum $\sum_1^n \int_a^x u_s(x)\,dx$, by $U_n(x)$; since $U(x) = \lim_{n=\infty} U_n(x)$, a value N of n can be found, such that, for $n \geqq N$,

$$|U(x) - U_n(x)|, \quad |U(x+h) - U_n(x+h)|$$

are both less than an arbitrarily chosen number δ, where $x+h$ is a fixed point in the neighbourhood $(x-\epsilon_1, x+\epsilon_2)$ of the point x already chosen. We now have

$$\left|\frac{U(x+h)-U(x)}{h}-\frac{U_n(x+h)-U_n(x)}{h}\right|<\frac{2\delta}{h}.$$

Now, since the interval $(x, x+h)$ contains no points at which the measure of non-uniform convergence of the series $\Sigma u(x)$ is indefinitely great, for a sufficiently great value of n,

$$\left|\int_x^{x+h} s_n(x)\,dx-\int_x^{x+h} s(x)\right|<\delta',$$

where δ' is arbitrarily chosen. Therefore, if $n \geqq N'$, where N' is some fixed integer, we have

$$\left|\frac{U_n(x+h)-U_n(x)}{h}-\frac{S(x+h)-S(x)}{h}\right|<\frac{\delta'}{h}.$$

From the two inequalities which have been obtained, we deduce that

$$\left|\frac{U(x+h)-U(x)}{h}-\frac{S(x+h)-S(x)}{h}\right|<\frac{2\delta+\delta'}{h}.$$

Since δ, δ' are arbitrarily small, and independent of x and h, we have

$$\frac{U(x+h)-U(x)}{h}=\frac{S(x+h)-S(x)}{h};$$

and this holds for any point x which does not belong to G, and for any point $x+h$, in a neighbourhood of G which does not contain points of G. It follows that any one of the four derivatives $D^+U(x)$, $D_+U(x)$, $D^-U(x)$, $D_-U(x)$, at x, is equal to the corresponding one of the four derivatives of $S(x)$. Since one of the four derivatives of the two functions $S(x)$, $U(x)$ is such that its value is the same for the two continuous functions at all points except at those of the enumerable closed set G, it follows (§ 206) that the two functions differ only by a constant; and since both vanish at $x=a$, they must be everywhere equal. It has thus been shewn that*:—

If the series $\Sigma u_n(x)$ converges to $s(x)$ in the interval (a, b), and if the integral $\int_a^x s(x)\,dx$ have everywhere a definite finite value, and $\sum_1^\infty \int_a^x u_n(x)\,dx$ have everywhere a definite finite value, and be a continuous function, it is a sufficient condition of the equality of the two, that the set of points at which the measure of non-uniform convergence is indefinitely great, should be an enumerable set.

When the set G is not enumerable, it contains a perfect component; and in that case the sum of the integrals of the terms of the series is not

* This theorem was given by Osgood, *American Journal of Math.*, vol. XIX, in the case in which the terms and the sum of the series are continuous. The general theorem was given by Arzelà, *Mem. di Bologna*, ser. 5, vol. VIII, 1900.

necessarily equal to the integral of the sum, even when both exist and the condition of continuity of $\sum_1^\infty \int_a^x u_n(x)\,dx$ is satisfied.

It will be observed that, in accordance with the theorems which have been demonstrated above, the term by term integration of a series may fail to give the integral of the sum, either (1) when the set G contains a finite, or enumerable, set of points, but the condition that the sum of the series $\Sigma \int_a^x u(x)\,dx$ is a continuous function of x is not satisfied; or (2) when G contains a perfect component; or (3) when the condition that the convergence of the series $\Sigma u(x)$ is of the kind called uniform convergence by intervals in general, is not satisfied, so that $s(x)$ is not integrable in accordance with Riemann's definition. In case (3), the term by term integration may, however, give the Lebesgue integral of $s(x)$.

386. We have hitherto assumed that the terms $u_n(x)$, of the series $\Sigma u_n(x)$, are all limited in the interval (a, b), and that the same holds as regards the sum-function $s(x)$. It is however possible, under certain restrictions, to remove these conditions. We shall assume that the functions $s_n(x)$, and therefore also the functions $S_n(x)$, are not necessarily all limited in the interval (a, b). In this case there may be values of x for which the series $\Sigma u_n(x)$ is divergent; and such points will be regarded as points of infinite discontinuity of $s(x)$, although $s(x)$ is not properly defined at such points. Let us assume that there is an enumerable closed set of points G, in (a, b), such that, in any interval (α, β) which contains, in its interior and at its ends, no point of G, the condition, that $|s_n(x)|$ is less than some fixed number, is satisfied for every value of n, and for the whole interval (α, β). Let us further assume that all the functions $u_n(x)$ possess improper, or proper, integrals in (a, b), and that the series $\sum_{n=1}^\infty \int_a^x u_n(x)\,dx$ is convergent for all values of x, and that its sum $U(x)$ is a continuous function of x, for the whole interval (a, b), their ends being included. Also let it be assumed that $s(x)$ has an improper integral in (a, b); then the function $\int_a^x s(x)\,dx = S(x)$ is a continuous function of x. The enumerable set G contains every point of divergence of the given series, and also every point of non-uniform convergence of which the measure is indefinitely great. It is now clear that, with these assumptions, the proof of the theorem in § 385 is applicable, without modification, to establish the legitimacy of term by term integration of the series $\Sigma u_n(x)$. We obtain therefore the following theorem:—

If the series $\Sigma u_n(x)$ converges to the function $s(x)$ at every point which does not belong to a reducible set of points G, and the functions $s_n(x)$, although not necessarily limited in (a, b), satisfy the condition that, in any *interval* (α, β) *which contains in its interior and at its ends no point of* G, $|s_n(x)|$ *is less than some fixed finite number, for every value of* n *and* x; *and if further* $\int_a^b u_n(x)\,dx$ *exist as an improper or proper integral, for every value of* n, *and the series* $\sum_{n=1}^{\infty}\int_a^x u_n(x)\,dx$, *for* $a \leqq x \leqq b$, *is convergent and represents a continuous function of* x; *and if* $s(x)$ *have an improper integral in* (a, b), *then the theorem*

$$\int_a^b s(x)\,dx = \lim_{n=\infty}\int_a^b s_n(x)\,dx$$

holds, and thus term by term integration is applicable to the series.

387. Lastly, the case will be considered in which the interval (a, b) is unlimited; we may assume that b has the improper value ∞. Let us suppose that term by term integration is applicable for every finite interval (a, C); and thus that

$$\lim_{n=\infty}\int_a^C s_n(x)\,dx = \int_a^C s(x)\,dx,$$

the integrals being either proper ones, or improper ones, subject to the conditions of the theorem in § 386. It follows that, if $C' > C$,

$$\lim_{n=\infty}\int_C^{C'} s_n(x)\,dx = \int_C^{C'} s(x)\,dx.$$

Let us now assume that, if ϵ be an arbitrarily chosen positive number, an integer n_1, and a number $C > a$, can be determined, such that

$$\left|\int_C^{C'} s_n(x)\,dx\right| < \epsilon,$$

for every value of $C' > C$, and for every value of $n \geqq n_1$. It then follows that $\left|\int_C^{C'} s(x)\,dx\right| \leqq \epsilon$; and since ϵ is arbitrary, it follows that $\int_a^{\infty} s(x)\,dx$ is convergent. We assume that all the integrals $\int_a^{\infty} s_n(x)\,dx$ exist. We have now

$$\left|\int_a^{\infty} s(x)\,dx - \int_a^{\infty} s_n(x)\,dx\right| \leqq \left|\int_a^C s(x)\,dx - \int_a^C s_n(x)\,dx\right| + \left|\int_C^{\infty} s(x)\,dx\right| + \left|\int_C^{\infty} s_n(x)\,dx\right|;$$

and by taking a sufficiently great value of $n \geqq n_1$, and a sufficiently great value of C, the expression on the right-hand side is $\leqq 3\epsilon$. It thus appears that

$$\int_a^{\infty} s(x)\,dx = \lim_{n=\infty}\int_a^{\infty} s_n(x)\,dx;$$

and therefore the following theorem has been established:—

If the series $\sum_{n=1}^{\infty} u_n(x)$ *have as its sum-function* $s(x)$, *in the sense previously defined, then if* $\int_a^{\infty} u_n(x)\,dx$ *exists for each value of* n, *and in every interval* (a, C) *the condition that* $\int_a^C s(x)\,dx = \sum_{n=1}^{\infty}\int_a^C u_n(x)\,dx$, *be satisfied; and further, if, corresponding to an arbitrarily chosen* ϵ, *an integer* n_1, *and a number* $C > a$ *can be so determined that* $\left|\int_C^{C'} s_n(x)\,dx\right| < \epsilon$, *for every value of* $C' > C$, *and for all values of* $n \geqq n_1$, *then the integral* $\int_a^{\infty} s(x)\,dx$ *exists, and is equal to* $\sum_{n=1}^{\infty}\int_a^{\infty} u_n(x)\,dx$.

It may also be shewn that, *on the assumption that the condition* $\int_a^C s(x)\,dx = \sum_{n=1}^{\infty}\int_a^C u_n(x)\,dx$ *holds for every value of* $C > a$, *then provided* $\sum_{n=1}^{\infty}\int_a^{\infty} u_n(x)\,dx$ *be convergent, and that* $\sum_{n=1}^{\infty}\int_a^C u_n(x)\,dx$ *converge to the value of* $\sum_{n=1}^{\infty}\int_a^{\infty} u_n(x)\,dx$, *when* C *is indefinitely increased, it follows that the integral* $\int_a^{\infty} s(x)\,dx$ *exists, and is equal to* $\sum_{n=1}^{\infty}\int_a^{\infty} u_n(x)\,dx$.

For, on the assumption that $\sum_{n=1}^{\infty}\int_a^C u_n(x)\,dx$ converges to a definite limit, as C is indefinitely increased, we see that, if ϵ be fixed, C may be so chosen that

$$\left|\sum_{n=1}^{\infty}\int_C^{C'} u_n(x)\,dx\right| < \epsilon,$$

for $C' > C$; and from this it follows that

$$\left|\int_C^{C'} s(x)\,dx\right| < \epsilon,$$

for $C' > C$; and since ϵ is arbitrary, it follows that $\int_a^{\infty} s(x)\,dx$ exists.

Also since
$$\int_a^C s(x)\,dx = \sum_{n=1}^{\infty}\int_a^C u_n(x)\,dx,$$
we see that $\int_a^{\infty} s(x)\,dx$ is equal to the limit to which

$$\sum_{n=1}^{\infty}\int_a^C u_n(x)\,dx$$

converges when C is indefinitely increased; and this limit is

$$\sum_{n=1}^{\infty}\int_a^{\infty} u_n(x)\,dx.$$

Therefore the theorem is established.

388. When $s(x)$ is the sum-function of a series $u_1(x)+u_2(x)+\dots$ in an interval (a, b), it is frequently desirable to know whether, for a function $F(x)$, defined for the interval (a, b), the series

$$\sum_{n=1}^{\infty}\int_a^b F(x)\,u_n(x)\,dx$$

converges to the limit $$\int_a^b F(x)\,s(x)\,dx.$$

If we assume that $s(x)$, $F(x)$ are limited integrable functions, it is clearly sufficient for the validity of the term by term integration, after multiplication by $F(x)$, that $|s_n(x)|$ should be less than some fixed finite number, for all values of x and n. For it then follows that $|F(x)\,s_n(x)|$ is less than a fixed finite number, for all values of n and x, and the result then follows by applying the theorem of § 383 to the series $\Sigma F(x)\,u_n(x)$, since $F(x)\,s(x)$ is an integrable function.

Again, *it is sufficient that $|s_n(x)|$ should be less than a fixed number C, for all values of x and n, and that $F(x)$ should have an absolutely convergent improper integral in (a, b).*

We may choose the integer n so that the measure of the set H, of those points at which

$$|s(x)-s_n(x)|>\epsilon,$$

is arbitrarily small, say η; we have then

$$\int_a^b F(x)\{s(x)-s_n(x)\}\,dx<\epsilon\int_a^b|F(x)|\,dx+2C\int_{\{H\}}|F(x)|\,dx,$$

and since ϵ and η are arbitrarily small, it follows that

$$\int_a^b F(x)\,s(x)\,dx=\lim_{n=\infty}\int_a^b F(x)\,s_n(x)\,dx;$$

and therefore the sufficiency of the criterion is established.

EXAMPLES.

1. Let $s_n(x)=nxe^{-nx^2}$, when n is odd, and $=0$, when n is even. In this case the series is simply-uniformly convergent; the sum $s(x)$ is the continuous function 0.

$$\int_0^x s_n(x)=\tfrac{1}{2}(1-e^{-nx^2}),\text{ or }0,$$

according as n is odd or even; thus

$$\lim_{n=\infty}\int_0^x s_n(x)\,dx$$

has no definite value, but $$\int_0^x s(x)=0.$$

The term by term integration fails in this case, because there is one point $x=0$, at which the measure of non-uniform convergence is indefinitely great, as may be seen from

$$R_n\left(\frac{1}{\sqrt{n}}\right)=-\sqrt{n}e^{-1},\ (n \text{ odd}),$$

the limit $\lim\int_0^x s_n(x)\,dx$ not being a continuous function of x.

2. Let $s_n(x)=2n^2xe^{-n^2x^2}$, then $s(x)=0$; at the point $x=0$, there is a point of indefinitely great measure of non-uniform convergence, since

$$s_n\left(\frac{1}{n}\right)=2ne^{-1},$$

$$\int_{x_0}^x s_n(x)\,dx=e^{-n^2x_0^2}-e^{-n^2x^2},\quad x_0<0.$$

If x be different from zero, $\lim\limits_{n=\infty}\int_{x_0}^x s_n(x)\,dx=0$, but at $x=0$ the limit is -1; thus, in any interval which contains the point 0, the function $\lim\limits_{n=\infty}\int_{x_0}^x s_n(x)\,dx$ is discontinuous, and therefore cannot equal $\int_{x_0}^x s(x)\,dx$, which is zero.

3. Let $$u_n(x)=\frac{x^{n-1}}{(n-1)!}+n^2xe^{-n^2x^2}-(n-1)^2xe^{-(n-1)^2x^2};$$

we find $s(0)=1$, and $s(x)=e^x$, for $|x|>0$.

We have $$\int_0^x s(x)\,dx=e^x-1.\quad \text{Also } \lim_{n=\infty}\int_0^x s_n(x)\,dx$$

is discontinuous at the point $x=0$, which is a point at which the measure of non-uniform convergence is infinite; it converges to $e^x-\frac{1}{2}$ if $x>0$, and to zero if $x=0$.

4. Let $$u_n(x)=\frac{k_n\phi_n'(x)}{1+\{\phi_n(x)\}^2}-\frac{k_{n+1}\phi_{n+1}'(x)}{1+\{\phi_{n+1}(x)\}^2}$$

where k_n is a function of n, and $\phi_n(x)$, $\phi_n'(x)$ are finite and continuous in the interval (a, b), and vanish for $x=a$. Further let it be assumed that $\phi_n(x)$, $\phi_n'(x)$ increase indefinitely with n, for every value of x except a, but so that $\lim\limits_{n=\infty} u_n(x)$ is zero.

We have $$\int_a^x s_n(x)\,dx=-k_{n+1}\tan^{-1}\{\phi_{n+1}(x)\}+k_1\tan^{-1}\{\phi_1(x)\},$$

$$\int_a^x s(x)\,dx=k_1\tan^{-1}\{\phi_1(x)\};$$

these are not identical unless $\quad k_{n+1}\tan^{-1}\{\phi_{n+1}(x)\}$

has the limit zero. If $\quad \phi_n(x)=h_n(x-a)^2,$

where h_n is positive and increases indefinitely with n, we have

$$\lim k_{n+1}\tan^{-1}\{\phi_{n+1}(x)\}=\tfrac{1}{2}\pi\lim k_{n+1}.$$

Hence, if $\lim k_{n+1}$ have a finite value, the two expressions have different finite values; if k_{n+1} increases indefinitely with n, the series of integrals of the terms of the series $\Sigma u_n(x)$ diverges. The series of integrals has in this case a point of discontinuity at $x=a$; we find that

$$s_n\left(a+\frac{1}{\sqrt{h_{n+1}}}\right)=\frac{2k_1h_1}{h_1^2+h_{n+1}^2}h_{n+1}^{\frac{3}{2}}-k_{n+1}h_{n+1}^{\frac{1}{2}},$$

and this increases indefinitely as n increases, and thus the point a is a point at which the measure of non-uniform convergence is indefinitely great.

5. Let
$$u_n(x)=\frac{2k_n h_n(x-a)}{1+h_n(x-a)^2}-\frac{2k_{n+1}h_{n+1}(x-a)}{1+h_{n+1}(x-a)^2},$$
where h_n increases indefinitely with n, and $k_n=\frac{c}{(\log h_n)^\beta}$.

In this case $s_n\left(a+\frac{1}{\sqrt{h_{n+1}}}\right)$ increases indefinitely with n, and thus a is a point of infinite measure of non-uniform convergence.

$$\lim_{n=\infty}\int_a^x s_n(x)\,dx=k_1\log\{1+h_1(x-a)^2\}-\lim_{n=\infty}k_{n+1}\log h_{n+1},\quad x>a,$$

and
$$\lim_{n=\infty}\int_a^x s_n(x)\,dx=0,\ \text{when } x=a;$$

also
$$\int_a^x s(x)\,dx=k_1\log\{1+h_1(x-a)^2\}.$$

If $\beta\leqq 1$, $\lim\limits_{n=\infty}k_{n+1}\log h_{n+1}$ is not zero, hence the term by term integration fails; but if $\beta>1$, this limit is zero, and the integral of $s(x)$ is equal to the sum of the series of integrals, although in either case the point a is a point of infinite measure of non-uniform convergence.

6. Let Γ be a perfect set of points constructed as follows:—In the middle of the interval (0, 1) lay off an interval (1) of length $l_1=\lambda-\frac{1}{3}\lambda$, where λ is a positive number not greater than unity. In the middle of each of the free end intervals, lay off an interval (2), both of these intervals to be of the same length l_2, and such that the total length of the intervals (1), (2) is $l_1+2l_2=\lambda-\frac{1}{4}\lambda$. Proceeding in this manner, in the middle of the equal free intervals, after $n-1$ such steps, lay off an interval (n), all these intervals to be of the same length l_n, and such that the total length of all the intervals (1), (2), ..., (n) is
$$l_1+2l_2+2^2l_3+\ldots+2^{n-1}l_n=\lambda-\frac{\lambda}{n+2}.$$
When n is indefinitely increased, the set of end-points of the intervals, and the limiting points of these end-points, form the perfect set Γ. Let
$$\psi_n(x)=nxe^{-nx^2},\quad x\geqq 0;$$
then form the function
$$\begin{aligned}\phi_n(x,l)&=\frac{\pi}{l}\sin\frac{\pi x}{l}\cdot\psi_n\left(\cos\frac{\pi x}{l}\right),\quad 0\leqq x\leqq\tfrac{1}{2}l,\\ &=-\frac{\pi}{l}\sin\frac{\pi x}{l}\cdot\psi_n\left(\cos\frac{\pi x}{l}\right),\quad -\frac{l}{2}\leqq x\leqq 0,\\ &=0,\ \text{for all other values of } x.\end{aligned}$$

Let the middle points of the above intervals (n) be denoted by $a_1^{(n)}, a_2^{(n)}, \ldots, a^{(n)}_{2^{n-1}}$, and let $s_n(x)$ be defined by
$$\begin{aligned}s_n(x)=&\phi_n(x-a_1^{(1)},l_1)+\phi_n(x-a_1^{(2)},l_2)\\ &+\phi_n(x-a_1^{(3)},l_3)+\ldots+\phi_n(x-a_4^{(3)},l_3)\\ &+\ldots\\ &+\phi_n(x-a_1^{(n)},l_n)+\ldots+\phi_n(x-a^{(n)}_{2^{n-1}},l_n).\end{aligned}$$

$s_n(x)$ is continuous in (0, 1), and converges to 0 for every value of x; for, if x_0 be a point of any interval (i), at most one term in the expression for $s_n(x)$ is different from zero, and this term converges to zero. If x_0 does not lie in any interval (i), all the terms of $s_n(x)$ are zero. Every point of the perfect set Γ is a point of infinite measure of non-uniform convergence of the series of which $s_n(x)$ is the partial sum. In this case the series $\Sigma\int_a^x u_n(x)\,dx$ is uniformly convergent, and thus has a continuous sum, which does not however coincide with the value of $\int_a^x s(x)\,dx$.

We find that
$$\int_{-\frac{1}{2}l_j}^{\frac{1}{2}l_j} \phi_n(x-a_i^{(j)}, l_j)\,dx = 1 - e^{-n},$$
$$\int_0^x s_n(x)\,dx = \frac{p_n}{2^{n-1}}(1-e^{-n}),$$
where $p_n < 2^{n-1}$, is the number of the intervals (n) which fall within $(0, x)$.

It can now be shewn that $\lim\int_0^x s_n(x)\,dx$ is a continuous function of x which increases from 0 to 1 as x increases from 0 to 1, whereas $\int_0^x \lim s_n(x)dx = 0$, for every value of x.

If any perfect non-dense set of points G be given, and $a_{n,m}$ be the middle point of the complementary interval of length $l_{n,m}$, the function
$$s_n(x) = \sum_{i=1}^{i=n}\sum_{j=1}^{j=k_i} \phi_n(x - a_{i,j}, l_{i,j})$$
will have, at every point of G, an infinite measure of non-uniform convergence to its limit $s(x)$. The intervals $l_{i,j}$ are here arranged in enumerable order, so that if $\epsilon_1, \epsilon_2, \ldots, \epsilon_i, \ldots$ be a descending sequence of positive numbers which converges to zero, $l_{i,1}, l_{i,2}, \ldots, l_{i,k_i}$ are those of which the lengths are $\leqq \epsilon_{i-1}$ and $> \epsilon_i$.

THE FUNDAMENTAL THEOREM OF THE INTEGRAL CALCULUS FOR LEBESGUE INTEGRALS.

389. Let $\phi(x)$ be a continuous function defined for the interval (a, b), and suppose that $\phi(x)$ has at every point of the interval a differential coefficient $f(x)$; let us further assume that $f(x)$ is limited in (a, b). The function $f(x)$ is definable as the limit of a sequence of continuous functions $\frac{\phi(x+h)-\phi(x)}{h}$, where h has the values in a sequence of which the limit is zero. It follows that $f(x)$ is of class 1, unless it be continuous, and it is consequently a summable function; in fact the theorem of § 381 shews that $f(x)$ has a Lebesgue integral, since $\left|\frac{\phi(x+h)-\phi(x)}{h}\right|$ has a finite upper limit for all values of x and h. In accordance with the last theorem of § 384, we have

$$\int_a^x f(x)\,dx = \lim_{h=0} \int_a^x \frac{\phi(x+h)-\phi(x)}{h}\,dx$$

$$= \lim_{h=0} \frac{1}{h}\left\{\int_x^{x+h} \phi(x)\,dx - \int_a^{a+h} \phi(x)\,dx\right\}$$

$$= \phi(x) - \phi(a),$$

since $\phi(x)$ is everywhere continuous. We have therefore established the theorem :—

If $\phi(x)$ be a function which possesses a differential coefficient $f(x)$, limited in an interval (a, b), then $f(x)$ always possesses an integral $F(x)$, in an interval (a, x), which differs from $\phi(x)$ by a constant only.

This theorem corresponds to the theorem (B) of § 258. An example due to Volterra has been given in § 264 of a function which possesses a limited differential coefficient that is not integrable in accordance with Riemann's definition.

The above theorem shews that this differential coefficient possesses a Lebesgue integral. It thus appears that the part (B) of the fundamental theorem of the Integral Calculus, as stated in § 258, holds without limitation, if Lebesgue's definition be employed, so long as the differential coefficient is a limited function.

390. Lebesgue has established* the following general theorem :—

In order that one of the four derivatives of a function may be integrable, that derivative being supposed finite at every point, it is necessary and sufficient that the function be of limited total fluctuation. Its total variation is the integral of the absolute value of the derivative.

The indefinite integral of such a summable derivative is the function of which it is the derivative.

This theorem affords a solution of the problem of the determination of a function when either its differential coefficient, or one of its four derivatives, is a given function, for the case in which that given function is limited, or also when it is known that the function to be determined must be of limited total fluctuation. This problem has been already considered in § 264.

In order to prove these theorems, we observe that, if $\phi(x)$ be a continuous function, defined for an interval (a, b), the derivatives $D^+\phi(x)$, $D_+\phi(x)$, are the upper, and the lower, limits of indeterminacy of $\lim_{h=0} I(x, h)$,

* *Leçons sur l'intégration*, p. 123.

where $I(x, h)$ denotes the incrementary ratio $\frac{\phi(x+h)-\phi(x)}{h}$. It will first be proved that a sequence of positive values of h can be determined, such that the upper and lower limits of $I(x, h)$, for $h=0$, are for every value of x the same when h has the successive values of the numbers in this sequence as when h is not restricted to have such values. Let $h_1', h_2', \dots h_n', \dots$ be a sequence of diminishing positive numbers, converging to zero, and let $\epsilon_1, \epsilon_2, \dots \epsilon_n, \dots$ be another such sequence. Since $I(x, h)$ is continuous with respect to (x, h), for all values of x in (a, b), and for positive values of h greater than zero, it follows, from the uniform continuity of $I(x, h)$, that the interval (h'_{n+1}, h_n') can be divided into a definite number r_n of parts, such that $|I(x,h)-I(x,h')|<\epsilon_n$, for every value of x, provided h and h' both lie in one and the same part of the interval. Let this subdivision be made for each value of n, and let $h_1, h_2, h_3, \dots$ denote the end-points of all the parts of all the intervals (h'_{n+1}, h_n'). The sequence $h_1, h_2, h_3, \dots$ converges to zero, and it is a sequence such as satisfies the required condition; for we have $|I(x, h_m)-I(x,h)|<\epsilon_s$, provided $h_m \geqq h \geqq h_{m+1}$, the integer s being determinate, corresponding to each value of m. It follows that the upper and the lower limits of the sequence $I(x, h_1), I(x, h_2), \dots I(x, h_m) \dots$ are identical with those of any other sequence $I(x, h_1'), I(x, h_2'), \dots I(x, h'_m) \dots$, where $h_1 \geqq h_1' \geqq h_2$, $h_2 \geqq h_2' \geqq h_3, \dots$, and generally $h_m \geqq h'_m \geqq h_{m+1}$; and this is the case for every value of x in (a, b). Therefore the sequence $\{h_n\}$ has the required property. Next, it will be proved that $D^+\phi(x)$ is a measurable function, and that it is at most of the second class. Let $u_1(x), u_2(x), \dots u_n(x), \dots$ denote a sequence of continuous functions, defined for an interval (a, b); and let $\bar{u}(x)$, $\underline{u}(x)$ denote the upper and lower limits of indeterminacy $\overline{\lim}_{n=\infty} u_n(x)$, $\underline{\lim}_{n=\infty} u_n(x)$. Let $v_n(x)$ denote that function which, for each value of x, has the value of the greatest of the functions $u_1(x), u_2(x), \dots u_n(x)$, for that value of x. It is easily seen that the functions $v_n(x)$ are all continuous in (a, b). The functions $v_1(x), v_2(x), \dots v_n(x) \dots$ form a sequence, which for each value of x is non-diminishing; let $w_1(x)$ denote its limit. The function $w_1(x)$ is measurable, and at most of the first class. Let the function $w_2(x)$ be formed in the same manner as $w_1(x)$, by leaving out the function $u_1(x)$, and proceeding as before. The function $w_n(x)$ is formed by leaving out the first $n-1$ of the functions $u(x)$, and then proceeding in the same manner as that in which $w_1(x)$ was formed from the original sequence. The functions $w_1(x), w_2(x), \dots w_n(x) \dots$ form a non-increasing sequence of measurable functions, of the first class at most; their limit is $\bar{u}(x)$. It follows that $\bar{u}(x)$ is measurable, and of the second class at most. In a similar manner it can be shewn that $\underline{u}(x)$ has the same property. If we identify the functions $u_n(x)$ with the functions $I(x, h_n)$, where the sequence $\{h_n\}$ is formed as has been explained above, we see that the two derivatives $D^+\phi(x)$, $D_+\phi(x)$ are measurable, and of the second class at most. The derivatives $D^- \phi(x)$, $D_- \phi(x)$ clearly have the same property.

We now assume that, for each value of x in (a, b), the derivative $D^+ \phi(x)$ has a finite value. Let the unlimited interval $(-\infty, \infty)$ be divided into intervals $(a_\iota, a_{\iota+1})$, where the integer ι has all positive and negative values, and so that, for each value of ι, $a_{\iota+1} - a_\iota < \epsilon$, where ϵ is a fixed positive number. Let e_ι denote that set of points x in (a, b), for which

$$a_\iota < D^+ \phi(x) \leqq a_{\iota+1};$$

and arrange the sets e_ι in the order $e_0, e_1, e_{-1}, e_2, e_{-2}, \dots e_n, e_{-n}, \dots$.

Let $k_0, k_1, k_{-1}, k_2, k_{-2}, \dots, k_n, k_{-n} \dots$ be a sequence of positive numbers so chosen that the limiting sum of

$$k_0 | a_0 | + k_1 | a_1 | + k_{-1} | a_{-1} | + \dots + k_n | a_n | + k_{-n} | a_{-n} | + \dots$$

is less than ϵ. Let the set e_0 be enclosed in a set of intervals Δ_0, and the complementary set $C(e_0)$ in a set of intervals Δ_0', so that the measure of that set of intervals which is common to the sets Δ_0, Δ_0' does not exceed k_0. Enclose e_1 in a set of intervals Δ_1, and $C(e_0 + e_1)$ in a set of intervals Δ_1'; where the sets Δ_1 and Δ_1' are both interior to Δ_0', and have in common a set of intervals of measure not exceeding k_1. Proceeding in this manner, we enclose e_p, where p is positive or negative, in a set of intervals Δ_p, and $C(e_0 + e_1 + e_{-1} + \dots e_q)$, where e_q immediately precedes e_p, in a set Δ'_p, so that Δ_p, Δ'_p are both interior to Δ'_q, and have in common a set of intervals of measure not exceeding k_p. We have now $m(\Delta_p) - m(e_p) \leqq k_p$, and Δ_p has in common with all the other sets Δ, a set of intervals whose measure is less than k_p.

Since $\sum_p | a_p | m(\Delta_p) - \Sigma | a_p | m(e_p)$, where $p = 0, 1, -1, 2, -2 \dots$, is less than $\sum_p k_p | a_p |$, or than ϵ, we see that

$$\sum_p | a_p | m(\Delta_p), \ \sum_p | a_p | m(e_p)$$

are either both divergent or both convergent; and in the latter case the difference of their sums is less than ϵ. If $\int_a^b | D^+ \phi(x) | dx$ exists, we have

$$\int_a^b | D^+ \phi(x) | dx < (b-a)\epsilon + \sum_p | a_p | m(e_p) < (b-a-1)\epsilon + \sum_p | a_p | m(\Delta_p);$$

and the integral exists if $\sum_p | a_p | m(\Delta_p)$ is convergent. Similarly, the necessary and sufficient condition that $\int_a^b D^+ \phi(x)\, dx$ should exist is that $\Sigma a_p m(\Delta_p)$ should converge; and then the two differ by less than $(b-a-1)\epsilon$. Any point x in (a, b) belongs to one of the sets e_p; let δ_p be that interval of the set Δ_p which contains x. Let $(x, x+h)$ be the longest interval with h positive, contained in δ_p, which does not exceed ϵ, and is also such that

$$a_p \leqq I(x, h) \leqq a_{p+1} + \epsilon.$$

Starting from the point a, we define in this manner an interval (a, x_1); then we take the interval (x_1, x_2) corresponding to x_1, and so on. The point b will be reached, either as the end-point x_β of an interval $(x_{\beta-1}, x_\beta)$, where β is some number of the first, or of the second class, or else it is the limiting point of the end-points of a sequence of intervals. The value of the sum

$$\Sigma \mid \phi(x_{\alpha+1}) - \phi(x_\alpha) \mid$$

taken for all the intervals $(x_\alpha, x_{\alpha+1})$, by which the points a and b are joined, lies between $\Sigma \mid a_{p_\alpha} \mid \delta'_{p_\alpha} \pm 2\epsilon(b-a)$, where δ'_{p_α} is the interval $(x_\alpha, x_{\alpha+1})$, and p_α denotes the corresponding value of p. Also $\Sigma \mid a_{p_\alpha} \mid \delta'_{p_\alpha}$ differs from $\sum\limits_p \mid a_p \mid m(\Delta_p)$ by less than $\sum\limits_p k_p \mid a_p \mid$, or by less than ϵ; for those points of Δ_p which do not belong to intervals δ'_{p_α}, necessarily belong to one of the sets Δ_q, where q is different from p, and their measure accordingly does not exceed k_p. Therefore $\Sigma \mid \phi(x_{\alpha+1}) - \phi(x_\alpha) \mid$ lies between the two numbers

$$\sum_p \mid a_p \mid m(\Delta_p) - \epsilon' \pm 2\epsilon(b-a), \text{ where } \epsilon' < \epsilon;$$

and these numbers are finite if $\int_a^b \mid D^+ \phi(x) \mid dx$ exists.

It follows that the necessary and sufficient condition that $D^+ \phi(x)$ should be integrable is that the function $\phi(x)$ should be of limited total fluctuation, in which case its total variation in (a, b) is limited.

We have also

$$\phi(b) - \phi(a) = \Sigma \{\phi_{\alpha+1}(x) - \phi_\alpha(x)\} = \int_a^b D^+ \phi(x)\, dx,$$

in case $D^+ \phi(x)$ is integrable; the reasoning being the same as before, and remembering that the number ϵ is arbitrarily small. The theorem may be proved for the case of the other derivatives in a precisely similar manner. *If then $\phi(x)$ be of limited total fluctuation, and have its four derivatives finite at each point, we have*

$$\phi(b) - \phi(a) = \int_a^b D^+ \phi(x) dx = \int_a^b D_+ \phi(x)\, dx = \int_a^b D^- \phi(x)\, dx = \int_a^b D_- \phi(x)\, dx.$$

391. The following theorem, also due to Lebesgue, will now be established :—

A function $\phi(x)$, of limited total fluctuation in (a, b), and of which one of the four derivatives is limited, has a differential coefficient $\phi'(x)$ at every point of (a, b), with the exception of points belonging to a set of measure zero. The theorem also holds in case the derivative has points of infinite discontinuity belonging to a closed set of points, of zero content.

Let e be a measurable set of points in (a, b), and let $e(x)$ denote the part of e in the interval (a, x). It will then be shewn that, $me(x)$ denoting the measure of $e(x)$, the function $me(x)$ has a differential coefficient equal to 1 at a set of points contained in e, and of measure $m(e)$; and that it has a differential coefficient equal to 0 at a set of points not belonging to e, and of measure $b-a-m(e)$. Thus the set of points at which $me(x)$ has no differential coefficient is of measure zero. All the points of e may be enclosed in the interiors of the intervals of sets $D_1, D_2, \ldots D_n, \ldots$, such that $m(D_n)$ converges to $m(e)$, as n is increased indefinitely. Let $E_n(x)$ denote the part of D_n which is in the interval (a, x); we then have

$$D^+ m\, E_n(x) \geqq D^+ me(x) \geqq 0.$$

But $D^+ m\, E_n(x) = 1$, at all interior points of D_n; and it follows that $D^+ m\, E_n(x) = 0$ at all points not interior to D_n, except at points of a set of measure zero; for otherwise

$$\int_a^b D^+ m\, E_n(x)\, dx$$

would exceed mD_n. As this holds for every value of n, it follows from the above inequality that $D^+ me(x)$ must be zero at every point not belonging either to e, or to a certain set of zero measure. Since $D^+ me(x) = 0$, at a set of points of measure $b-a-m(e)$, and since

$$\int_a^b D^+ me(x)\, dx = m(e),$$

it follows that $D^+ me(x)$ is equal to unity at a set of points of measure $m(e)$, belonging to the inner limiting set defined by the sequence $\{D_n\}$. Also this inner limiting set has the measure $m(e)$; therefore $D^+ me(x) = 1$, at a set of points all belonging to e, and of measure $m(e)$. A similar theorem can be established for each of the other derivatives of $me(x)$. It then follows that $me(x)$ has a differential coefficient equal to 1, at a set of points of measure $m(e)$, belonging to e, and also a differential coefficient equal to 0, at a set of points of measure $b-a-m(e)$, belonging to $C(e)$.

Next, let $f(x)$ be a limited summable function defined in (a, b), and of which L, U are the lower and the upper limits. Divide the interval (L, U) into parts

$$(a_0, a_1),\ (a_1, a_2) \ldots (a_{\iota-1}, a_\iota) \ldots (a_{n-1}, a_n), \text{ where } a_0 = L,\ a_n = U;$$

and where $a_\iota - a_{\iota-1}$ does not exceed ϵ, for any value of ι. Let $\phi_1(x)$, $\phi_2(x)$ be two functions defined as follows:—For each value of x such that

$$a_\iota \leqq f(x) < a_{\iota+1}, \text{ let } \phi_1(x) = a_\iota;$$

and for each value of x such that $a_{\iota-1} < f(x) \leqq a_\iota$, let $\phi_2(x) = a_\iota$. Thus $\phi_1(x)$, $\phi_2(x)$ are defined for the whole interval (a, b), each of them having only a finite set of values. Let

$$F(x) = \int_a^x f(x)\, dx,\quad F_1(x) = \int_a^x \phi_1(x)\, dx,\quad F_2(x) = \int_a^x \phi_2(x)\, dx;$$

we then have $D^+ F_1(x) \leqq D^+ F(x) \leqq D^+ F_2(x)$, at each point x in (a, b). Also, we have

$$F_1(x) = a_0\, me_0(x) + a_1\, me_1(x) + \ldots + a_n\, me_n(x),$$

where $e_\iota(x)$ is the part, contained in (a, x), of the set e_ι of points of (a, b) at which $a_\iota \leqq f(x) < a_{\iota+1}$. The function $F_1(x)$ has a differential coefficient equal to $\phi_1(x)$, at all points of (a, b) not belonging to a set of measure zero. For $me_\iota(x)$ has a differential coefficient at every point not belonging to a set of zero measure, and this holds for every value of ι; therefore $F_1(x)$ has a differential coefficient at every point not belonging to a set of measure zero. Also $\frac{d}{dx} me_\iota(x)$ is unity at all points of e_ι except at points of a set of zero measure, and is zero at all points of $C(e_\iota)$ except at points of a set of zero measure; and this holds for each value of ι. A similar result holds for the function $F_2(x)$.

At any point x, we have

$$\frac{F_1(x+h) - F_1(x)}{h} < \frac{F(x+h) - F(x)}{h} < \frac{F_2(x+h) - F_2(x)}{h}$$

for positive or negative values of h. Therefore the four derivatives of $F(x)$ all lie between $\phi_1(x)$, and $\phi_2(x)$, at each point x which does not belong to that set of points of zero measure at which $F_1(x)$, $F_2(x)$ do not possess differential coefficients equal to $\phi_1(x)$, $\phi_2(x)$. Now $\phi_1(x)$, $\phi_2(x)$ differ from one another, and from $f(x)$, by not more than ϵ; therefore, at every point not belonging to a set of measure zero, the four derivatives of $F(x)$ do not differ from one another by more than ϵ. By taking a sequence of values of ϵ converging to zero, we then see that $F(x)$ has a differential coefficient equal to $f(x)$ at every point of (a, b) not belonging to a set of zero measure.

If the function $f(x)$ be unlimited, but summable and possessing a Lebesgue integral, the functions $F_1(x)$, $F_2(x)$ are each defined by series, infinite in both directions. The term by term differentiation of this series would then require justification. It will be, however, sufficient* for the present purpose to consider the case in which the points of infinite discontinuity of $f(x)$ form a closed set of zero content. By enclosing this set of points in a finite set of intervals of arbitrarily small sum, and applying the result obtained above to each of the complementary intervals in which $f(x)$ is limited, the theorem may be extended to the case of such a function $f(x)$.

* [illegible]besgue applies reasoning similar to that in the text to the case of any summable function, [illegible] 124. As each of his functions ϕ is the sum of an infinite number of functions ψ, when [illegible]), the process would then appear to require some further justification.

It has now been established that, *if* $f(x)$ *be a summable function which is either limited in* (a, b), *or has points of infinite discontinuity belonging to a closed set of points of zero content, then* $\int_a^x f(x)\,dx$ *has a differential coefficient equal to* $f(x)$, *at a set of points whose measure is equal to that of the whole interval* (a, b).

The function $\phi(x)$, being of limited total fluctuation, is equal to the integral $\int_a^b D^+\phi(x)\,dx$, the integrand being supposed finite at each point. It has now been shewn that, if the derivative $D^+\phi(x)$ is limited, or at most has indefinitely great values in the neighbourhoods of points of a closed set of zero content, then $\phi(x)$ has a differential coefficient equal to $D^+\phi(x)$ at each point of a set of measure $b-a$. At such a point the four derivatives are of course equal to one another. Therefore the theorem stated at the beginning of the present section has been established.

Lebesgue has also established* the following theorem :—

Every function with limited total fluctuation, and in particular, every monotone function, has a finite differential coefficient, except at the points of a set of which the measure is zero.

This differential coefficient is summable in the domain which consists of those points at which it exists and is finite, but the integral is not necessarily the given function, unless one of the four derivatives of the given function be everywhere finite.

392. Besides the Lebesgue integrals of unlimited functions, as defined in § 291, which integrals are necessarily absolutely convergent, there is a class of *non-absolutely convergent improper Lebesgue integrals*, which may be defined by extending Harnack's definition of improper integrals given in § 271. The extension consists in taking the integrals in the intervals η employed in § 271, to be Lebesgue integrals, and not necessarily Riemann integrals. In case the integral so defined, of an unlimited function, be absolutely convergent, it has been shewn in § 291, to be in agreement with the ordinary Lebesgue integral of the same function, as defined in § 291. If, however, the integrals taken through the set of intervals $\{\eta\}$ have a limit, but the limit of the integrals of the absolute values of the functions do not exist, we have then a non-absolutely convergent improper Lebesgue integral.

An example of such an integral is the following :—Let $f(x)=0$, for all rational values of x in $(0, 1)$; and for irrational values of x, let $f(x)=\frac{1}{x}\sin\frac{1}{x}$; then $\int_0^1 f(x)\,dx$ exists only as a non-absolutely convergent Lebesgue integral,

* *Leçons sur l'intégration*, pp. 123 and 128.

being defined as $\lim_{\epsilon=0}\int_\epsilon^1 f(x)\,dx$. The integral $\int_0^1 |f(x)|\,dx$ does not exist; for $\int_\epsilon^1 |f(x)|\,dx$ is not convergent, for $\epsilon=0$.

393. The theorem proved in § 282 that, if $\int_a^b f(x)\,dx$ exist as an improper integral, whether absolutely convergent or not, in accordance with Harnack's definition, then $\int_{a'}^{b'} f(x)\,dx$ exists, where $a \leqq a' \leqq b' \leqq b$, and that the convergence of this integral is uniform for all values of a' and b', is applicable without change to the case of improper Lebesgue integrals, whether absolutely convergent or not. The proof in § 282, is valid in this more general case.

Also the proof that

$$\int_a^b f(x)\,dx = \int_a^x f(x)\,dx + \int_x^b f(x)\,dx$$

given in § 282, is applicable without change.

It may be proved that, *for an improper Lebesgue integral,* $\int_a^x f(x)\,dx$ *is a continuous function of the upper limit* x.

Using the notation of § 282, we have

$$\left|\int_a^{x+h} f(x)\,dx - \int_a^{x+h} f_\delta(x)\,dx\right| < \tfrac{1}{3}\epsilon,$$

and

$$\left|\int_a^{x} f(x)\,dx - \int_a^{x} f_\delta(x)\,dx\right| < \tfrac{1}{3}\epsilon,$$

provided the set of intervals $\{\delta\}$ is properly chosen. We have also, since $f_\delta(x)$ is a limited function,

$$\left|\int_a^{x+h} f_\delta(x)dx - \int_a^{x} f_\delta(x)\,dx\right| < \tfrac{1}{3}\epsilon,$$

provided $|h|$ is less than some positive number η. It follows that

$$\left|\int_a^{x+h} f(x)\,dx - \int_a^{x} f(x)\,dx\right| < \epsilon, \text{ if } |h| < \eta.$$

Therefore, since ϵ is arbitrary, $\int_a^x f(x)\,dx$ is continuous.

It has been shewn in § 390, that if $\phi(x)$ have limited total fluctuation in (a, b), and if $D\phi(x)$ be everywhere finite, then

$$\phi(b) - \phi(a) = \int_a^b D\phi(x)\,dx.$$

Now let $D\phi(x)$ be indefinitely great at points of a reducible set G; then if (a', x) be interior to one of the complementary intervals of the set G, and $\phi(x)$ have limited total fluctuation in the complementary interval, we have

$$\phi(x) - \phi(a') = \int_{a'}^{x} D\phi(x)dx;$$

and if $D\phi(x)$ have a Lebesgue integral, or a non-absolutely convergent improper Lebesgue integral in (a, b), we have

$$\int_a^x D\phi(x)\,dx - \int_a^{a'} D\phi(x)\,dx = \phi(x) - \phi(a').$$

Therefore $\int_a^x D\phi(x)\,dx - \phi(x)$ is constant throughout the interior of the complementary interval considered. The function $\phi(x)$ being continuous, this difference is continuous throughout (a, b); and since G is reducible, it follows from the theorem of § 206, that $\int_a^x D\phi(x)\,dx - \phi(x)$ is constant throughout (a, b), and therefore $= -\phi(a)$. The following extension of the theorem of § 390, has therefore been established:—

If $\phi(x)$ be a continuous function such that $D\phi(x)$ is finite at every point of (a, b) which does not belong to a reducible set G, and if $\int D\phi(x)\,dx$ exist as a Lebesgue integral, or as an improper non-absolutely convergent Lebesgue integral, then

$$\int_a^x D\phi(x)\,dx = \phi(x) - \phi(a)$$

for every point x in (a, b). $D\phi(x)$ denotes any one of the four derivatives of $\phi(x)$.

If the set G were not reducible, but contained a perfect component, then $\int_a^x D\phi(x)\,dx$ would in general differ from $\phi(x)$ by a function with an everywhere-dense set of lines of invariability.

INTEGRATION BY PARTS FOR LEBESGUE INTEGRALS.

394. If u, v be two continuous functions with limited total fluctuation in (a, b), then the product uv has the same property. For if $u = u_1 - u_2$, $v = v_1 - v_2$, where u_1, u_2, v_1, v_2 are monotone non-diminishing functions, then $u_1v_1 + u_2v_2$, $u_1v_2 + u_2v_1$ have the same property, and therefore $(u_1 - u_2)(v_1 - v_2)$ is of limited total fluctuation.

Let us assume that u, v both have limited derivatives in (a, b); then in accordance with the theorem of § 391, $\frac{du}{dx}, \frac{dv}{dx}$ both exist at all points of (a, b), except points of a set with zero measure.

We have $$\frac{d(uv)}{dx} = v\frac{du}{dx} + u\frac{dv}{dx},$$

at each point of the set E of points where the differential coefficients exist. We have then

$$\int_E \frac{d(uv)}{dx}\,dx = \int_E v\frac{du}{dx}\,dx + \int_E u\frac{dv}{dx}\,dx.$$

Also $$\int_E v\frac{du}{dx}\,dx = \int_a^b v\,Du\,dx, \quad \int_E u\frac{dv}{dx}\,dx = \int_a^b u\,Dv\,dx,$$

since Du, Dv are limited, and $m(E) = b - a$. Also

$$\int_E \frac{d(uv)}{dx}\,dx = \Big[uv\Big]_a^b;$$

therefore $$\int_a^b u\,Dv\,dx = \Big[uv\Big]_a^b - \int_a^b v\,Du\,dx.$$

Now let $$u = \int_\alpha^x U\,dx, \quad v = \int_\beta^x V\,dx,$$

where U and V are limited in (a, b); then U only differs from Du at points of a set of zero measure, and V differs from Dv only in the same manner. We thus obtain the formula for integration by parts,

$$\int_a^b V\left(\int_\alpha^x U\,dx\right)dx = \left[\left(\int_\alpha^x U\,dx\right)\left(\int_\beta^x V\,dx\right)\right] - \int_a^b U\left(\int_\beta^x V\,dx\right)dx,$$

where α, β are arbitrarily fixed points in the interval (a, b).

If Du, Dv be not limited, but have points of infinite discontinuity which belong to an enumerable closed set G, let $\chi(x)$ denote

$$\Big[uv\Big]_a^x - \int_a^x u\,Dv\,dx - \int_a^x v\,Du\,dx.$$

The function $\chi(x)$ is constant in any interval contained in an interval complementary to G. The functions u, v being continuous in (a, b), the function $\chi(x)$ is continuous in (a, b), and therefore, since G is enumerable, $\chi(x)$ is constant throughout (a, b); and it is zero, since it vanishes at the point a.

Therefore we have

$$\int_a^b u\,Dv = \Big[uv\Big]_a^b - \int_a^b v\,Du\,dx,$$

where u, v are continuous, and Du, Dv have points of infinite discontinuity belonging to a reducible set of points.

THE DIFFERENTIATION OF SERIES.

395. If $s(x)$ denote the sum-function of a series $u_1(x)+u_2(x)+\ldots$, and it be assumed that, either at a particular point x, or in a continuous interval, all the terms $u_1(x)$, $u_2(x)$, ... are continuous and differentiable, it is a subject for investigation under what conditions $s(x)$ possesses a differential coefficient which is the limit of the sum $u_1'(x)+u_2'(x)+\ldots$, the series of which the terms are the differential coefficients of the original series. It may happen that (1) $s(x)$ possesses no differential coefficient, or (2) that the series $u_1'(x)+u_2'(x)+\ldots$ is not convergent, or both (1) and (2) may happen simultaneously, or (3) that $s'(x)$ exists and the series is also convergent, but that its limiting sum is not $s'(x)$.

Writing $s(x)=s_n(x)+R_n(x)$, we have, at any point of convergence of the series, $\lim\limits_{n=\infty} R_n(x)=0$; further we have

$$\frac{s(x+h)-s(x)}{h}=\frac{s_n(x+h)-s_n(x)}{h}+\frac{R_n(x+h)-R_n(x)}{h}.$$

On the hypothesis that all the terms of the series have finite differential coefficients at the point x, we have $\lim\limits_{h=0}\frac{s_n(x+h)-s_n(x)}{h}=s_n'(x)$; if then $R_n(x)$ possesses a differential coefficient, so also does $s(x)$. If $R_n'(x)$ exists, and converges to the limit zero, when n is indefinitely increased, we have

$$s'(x)=\lim_{n=\infty} s_n'(x)=\lim_{n=\infty}\{u_1'(x)+u_2'(x)+\ldots+u_n'(x)\}.$$

In case $R_n'(x)$ either does not exist, or exists but does not converge to the limit zero, when n is increased indefinitely, the term by term differentiation of the series is inapplicable.

396. Let us assume that, in a given interval (a, b), the terms of the convergent series $u_1(x)+u_2(x)+\ldots+u_n(x)+\ldots$ are differentiable, and that their differential coefficients are everywhere finite, and are integrable in (a, b), and in case they are unlimited have a reducible set of points of infinite discontinuity, so that $\int_a^x u_n'(x)\,dx = u_n(x)-u_n(a)$. The integral may be either a Riemann integral, or an improper one in accordance with Harnack's definition, or a Lebesgue integral, or a non-absolutely convergent improper Lebesgue integral. Let it be further assumed that the series

$$u_1'(x)+u_2'(x)+\ldots+u_n'(x)+\ldots$$

is convergent everywhere in (a, b); then denoting the sum-function of this latter series by $\phi(x)$, we may apply the theorems in §§ 379–386, to obtain conditions that $\phi(x)$ possesses an integral $\int_a^x \phi(x)\,dx$, where $a \leqq x \leqq b$, and that the series

$$\{u_1(x)-u_1(a)\}+\{u_2(x)-u_2(a)\}+\ldots$$

converges to the sum $\int_a^x \phi(x)\,dx$. If these conditions are satisfied, we have $\int_a^x \phi(x)\,dx = s(x) - s(a)$; from which it follows that, at least at any point of continuity of $\phi(x)$, the differential coefficient $s'(x)$ exists, and is the sum of the series

$$u_1'(x) + u_2'(x) + \ldots + u_n'(x) + \ldots.$$

It follows from a theorem in § 383, that if the series $\Sigma u_n'(x)$ be uniformly convergent, the function $\phi(x)$ is integrable in (a, b), and that $\int_a^x \phi(x)\,dx = s(x) - s(a)$; and thus $s'(x) = \phi(x)$, at any point of continuity of $\phi(x)$. In particular, $\phi(x)$ is everywhere continuous if all the terms $u_n'(x)$ are continuous functions. We have therefore established the following theorems:—

If the series $\Sigma u_n(x)$ converge in (a, b), and the terms of the series $\Sigma u_n'(x)$ be all finite and continuous in (a, b), and the latter series converge uniformly, then $s'(x)$ exists, and is the sum of the series $\Sigma u_n'(x)$, at all points in (a, b).

If the series $\Sigma u_n(x)$ converge in (a, b), and the differential coefficients $u_n'(x)$ have all definite finite values everywhere in (a, b), and are integrable in the sense explained above, and the series $\Sigma u_n'(x)$ be uniformly convergent; then at every point of continuity of the sum of the series $\Sigma u_n'(x)$, that sum is $s'(x)$.

It is a known theorem that the sum-function $\phi(x)$ of a uniformly convergent series of point-wise discontinuous functions is at most point-wise discontinuous; and in the present case the points of discontinuity form a set of zero measure, provided the integrals of the terms $u_n'(x)$ are Riemann integrals.

Exactly similar theorems hold for derivatives of the terms $u_n(x)$, on one side.

The condition of uniform convergence contained in these theorems is a sufficient, but not a necessary, condition for the validity of the process of term by term differentiation.

A less stringent, but sufficient, condition would, in accordance with the last theorem of § 383, be obtained by replacing the condition that $\Sigma u_n'(x)$ should converge uniformly in (a, b) by the condition that its convergence should be simply uniform*.

Still wider conditions for the validity of the process are obtained by applying the theorems of §§ 384, 385. We thus obtain the following theorems:—

If the series $\Sigma u_n(x)$ converge in (a, b), and the differential coefficients $u_n'(x)$ everywhere exist, and are limited, and the series $\Sigma u_n'(x)$ be every-

* See Bendixson, "Sur la convergence uniforme des séries," *Stockholm Öfv.* vol. LIV, 1897.

where convergent; and if further $\left|\sum_1^n u_n'(x)\right|$ *be, for every value of n and x, less than some fixed positive number, then* $\sum_1^\infty u_n'(x) = \frac{d}{dx}\sum_1^\infty u_n(x)$, *at least at every point of continuity of* $\sum_1^\infty u_n'(x)$.

If the series $\Sigma u_n(x)$ *converge in* (a, b), *and the functions* $u_n'(x)$ *everywhere exist, and are limited, then provided the set of points, in the neighbourhood of which the condition that* $\left|\sum_1^m u_n'(x)\right|$ *is for every n and x less than a fixed positive number is not satisfied, be an enumerable set, and if also* $\sum_1^\infty u(x)$ *be continuous in* (a, b), *then* $\sum_1^\infty u_n'(x) = \frac{d}{dx}\sum_1^\infty u_n(x)$, *at least at every point of continuity of* $\sum_1^\infty u_n'(x)$.

A particular case of this theorem is that in which all the terms $u_n'(x)$, and $\sum_1^\infty u_n'(x)$, are continuous. In the general case, it will be observed that, in accordance with the theorem of § 389, the terms $u_n'(x)$ are all integrable, possessing at least Lebesgue integrals. Also, in virtue of the theorem proved in § 391, the points at which the term by term differentiation does not hold, form at most a set of points of measure zero.

By employing the theorem established in § 386, the above theorems can be extended to the case in which the series $\sum_1^\infty u_n'(x)$ fails to converge at points belonging to a reducible set of points.

397. The condition of the validity of term by term differentiation of the convergent series $\Sigma u(x)$, at a particular point α of the domain of x, is identical with the condition that the two repeated limits of

$$\frac{s(\alpha + h, y) - s(\alpha, y)}{h}$$

for $h = 0$, $y = 0$, should exist, and have one and the same value. By applying the theorems of §§ 234, 235, which contain the necessary and sufficient conditions for the existence and equality of repeated limits of a function at a point, we obtain the following theorems:—

If the series $\Sigma u_n(x)$ *everywhere converge in a sufficiently small neighbourhood of a point* α, *and the differential coefficients* $u_n'(\alpha)$ *exist, and are finite, then the necessary and sufficient conditions that* $\frac{d}{dx}s(x)$ *at* $x = \alpha$, *may exist and be equal to* $\sum_1^\infty u_n'(\alpha)$ *are (1) that* $\Sigma u_n'(\alpha)$ *be convergent, and (2) that,* ϵ *being an arbitrarily chosen positive number, and* n_0 *an arbitrarily chosen positive integer,*

36—2

a number η, positive and > 0 can be found, and also a positive integer $n > n_0$, such that the condition $\left|\frac{R_n(\alpha+h) - R_n(\alpha)}{h}\right| < \epsilon$ *is satisfied for this value of n, and for every value of h such that $0 < |h| < \eta$, and for which $\alpha + h$ is interior to the given neighbourhood of α.*

If the series $\Sigma u_n(x)$ converge everywhere in a sufficiently small neighbourhood of a point α, and the differential coefficients $u_n'(\alpha)$ exist and are finite, then the necessary and sufficient condition that $\frac{d}{dx}s(x)$ at $x = \alpha$, may exist and be equal to $\sum_1^{\infty} u_n'(\alpha)$ is that, corresponding to any arbitrarily chosen positive number ϵ, an integer n_0 exists, such that corresponding to each integer $n > n_0$, a positive number η, in general dependent on n, can be found, such that the condition $\left|\frac{R_n(\alpha+h) - R_n(\alpha)}{h}\right| < \epsilon$ *is satisfied for every value of h such that $0 < |h| < \eta$, and for which $\alpha + h$ is interior to the given neighbourhood of α.*

It is clear from § 234, that *the uniform convergence of* $\frac{s_n(\alpha+h) - s_n(\alpha)}{h}$ *to the limit* $\frac{s(\alpha+h) - s(\alpha)}{h}$, *for all values of h, except 0, in a fixed interval $(-\delta, \delta')$ for h, is a sufficient condition that $s'(\alpha)$ exists, and that the series $\Sigma u_n'(\alpha)$ converges to $s'(\alpha)$.*

398. The following theorem* is frequently more convenient than the theorems of § 397, for the purpose of ascertaining whether a function defined by a convergent series of functions is differentiable or not.

If the series $\Sigma u_n(x)$ converge in (a, b), and the differential coefficients $u_n'(\alpha)$ exist, and are finite, then the necessary and sufficient conditions that $\frac{d}{dx}s(x)$ may exist at $x = \alpha$, and be the sum of the series $\Sigma u_n'(\alpha)$, are (1) that the series $\Sigma u_n'(\alpha)$ be convergent, and (2) that, corresponding to an arbitrarily fixed positive number ϵ, and an arbitrarily fixed integer m', a positive number δ can be determined such that, for each value of h numerically less than δ, and for which $\alpha + h$ is in (a, b), an integer $m\,(> m')$, in general varying with h, can be found, for which the three numbers

$$\sum_{n=1}^{m}\left\{\frac{u_n(\alpha+h) - u_n(\alpha)}{h} - u_n'(\alpha)\right\},\quad \frac{R_m(\alpha+h)}{h},\quad \frac{R_m(\alpha)}{h}$$

are all numerically less than ϵ.

The convenience in application of this theorem arises from the fact that

* Dini, *Grundlagen*, p. 152.

it provides a test in which only a single value of h is employed. To prove that the conditions stated in the theorem are sufficient, we have

$$\frac{s(\alpha+h)-s(\alpha)}{h}-\sum_{n=1}^{\infty}u_n'(\alpha)=\sum_{1}^{m}\left\{\frac{u_n(\alpha+h)-u_n(\alpha)}{h}-u_n'(\alpha)\right\}+\frac{R_m(\alpha+h)}{h}-\frac{R_m(\alpha)}{h}-R_m'$$

where R_m' denotes the remainder, after m terms, of the series $\Sigma u_n'(\alpha)$. The number m' can be so chosen that $|R_n'|<\epsilon$, for $n \geqq m'$, since the series $\Sigma u_n'(\alpha)$ is convergent. If m be chosen $> m'$, and such that the second condition in the theorem is satisfied, we see that $\left|\frac{s(\alpha+h)-s(\alpha)}{h}-\sum_{n=1}^{\infty}u_n'(\alpha)\right|<4\epsilon$, provided $|h|<\delta$; and therefore $\lim \frac{s(\alpha+h)-s(\alpha)}{h}$ is $\sum_{n=1}^{\infty}u_n'(\alpha)$. Therefore the conditions are sufficient.

To shew that the conditions stated are necessary; it is clear that (1) must be satisfied, and therefore that m' can be determined so that $|R_m'|<\frac{1}{4}\epsilon$, if $m \geqq m'$. Moreover, a positive number δ can be determined such that $\frac{s(\alpha+h)-s(\alpha)}{h}-\sum_{n=1}^{\infty}u_n'(\alpha)$ is numerically less than $\frac{1}{4}\epsilon$, if $|h|<\delta$. Also since $\Sigma u_n(x)$ is convergent, for each value of h, a corresponding value of $m\,(\geqq m')$ exists, such that $\frac{R_m(\alpha+h)}{h}$, $\frac{R_m(\alpha)}{h}$ are each numerically $<\frac{1}{4}\epsilon$. It then follows that, for these values of h and m, the condition

$$\sum_{1}^{m}\left\{\frac{u_n(\alpha+h)-u_n(\alpha)}{h}-u_n'(\alpha)\right\}<\epsilon$$

is satisfied. Therefore the conditions in the theorem are necessary.

EXAMPLES.

1. Let $u_n(x)=\frac{1}{n}\sin nx$; the series $\Sigma u_n(x)$ converges everywhere in any interval, but the series $\Sigma \cos nx$ does not converge. The term by term differentiation of the given series is therefore inapplicable.

2. Let $u_n(x)=\frac{x^n}{n}-\frac{x^{n+1}}{n+1}$; the series $\Sigma u_n(x)$ converges to the sum-function $s(x)=x$, in the interval (0, 1). The series $\Sigma(x^{n-1}-x^n)$ converges to $s'(x)=1$, for all values of x in the interval (0, 1), except for $x=0$, when it converges to 0, which is not equal to $s'(0)$. The series $\Sigma(x^{n-1}-x^n)$ has the point $x=0$ for a point of non-uniform convergence, and thus the convergence is not uniform in the interval (0, 1).

3. The series $\sum_{n=1} b^n\cos(a^n x)$, where $0<b<1$, converges uniformly in any interval. The series $-\Sigma(ab)^n\sin(a^n x)$, for $ab>1$, is not convergent. It will be shewn later that the function defined by the given series is not differentiable for any value of x, provided ab exceeds a certain value.

REPEATED IMPROPER INTEGRALS.

399. If $f(x, y)$ be an unlimited function, defined in the fundamental rectangle bounded by $x=a$, $x=b$, $y=c$, $y=d$, it is an important case of the problem of the reversal of the order of repeated limits, to investigate conditions under which the repeated integrals

$$\int_a^b dx \int_c^d f(x, y)\, dy, \quad \int_c^d dy \int_a^b f(x, y)\, dx$$

when they both exist, have the same value. In the case in which the improper double integral $\int f(x, y)(dx\, dy)$ exists, it is also a matter for investigation whether, or under what conditions, this double integral can be replaced by one or other of the corresponding repeated integrals. These problems have been investigated* by de la Vallée-Poussin, who has obtained a number of results of a general character. An investigation† will be here given, in which the questions are considered in a still more general manner, free from some of the restrictions introduced by de la Vallée-Poussin.

The definition of a double integral will be taken to be that of § 321; that of a single integral will be taken to be that of Riemann, or that of de la Vallée-Poussin, in the case of an unlimited function.

An extension of the definition, given in § 379, of the regular convergence of a sequence of functions to a limiting function will be first given.

Let $\phi_1(x), \phi_2(x), \ldots \phi_n(x), \ldots$ be a sequence of functions defined for the interval (a, b). We shall suppose that, for each value of x, any one of these functions $\phi_n(x)$ has either a definite value, or is multiple-valued; and in the latter case it is regarded as indeterminate between limits of indeterminacy, either of which may be finite or infinite, of which the upper limit may be denoted by $\overline{\phi_n(x)}$, and the lower limit by $\underline{\phi_n(x)}$. For any value of x for which $\phi_n(x)$ is determinate, we have $\overline{\phi_n(x)} = \underline{\phi_n(x)}$. When either $\overline{\phi_n(x)}$ or $\underline{\phi_n(x)}$ is to be taken indifferently, we may use the notation $\overline{\underline{\phi_n(x)}}$.

The consideration of a function $\phi_n(x)$ which, for a particular set of values of x, is indeterminate, as a single function, involves an extension of Dirichlet's definition of a function which is justified by its convenience for use in investigations such as the present one. This extension, which has been already referred to in § 166, is convenient when the functional value $\phi_n(x)$

* His investigations are contained in three memoirs, the first in the *Annales de la Société scientifique de Bruxelles*, vol. XVII; the second in *Liouville's Journal*, ser. 4, vol. VIII, 1892; and the third in *Liouville's Journal*, ser. 5, vol. V.

† See also Hobson, *Proc. Lond. Math. Soc.*, ser. 2, vol. IV, 1906.

at a point x is defined by means of a limit, say $\phi_n(x) = \lim_{m=\infty} \psi_n(x, m)$, which may be such that, for a particular value of x, $\lim_{m=\infty} \psi_n(x, m)$ has no single value, but may be multiple-valued between limits $\overline{\phi_n(x)}$, $\underline{\phi_n(x)}$. The function $\phi_n(x)$, for such a value of x, may be capable of having a finite number, or an infinite number, of values, and possibly of having all values between $\overline{\phi_n(x)}$ and $\underline{\phi_n(x)}$; but in the application of the theory we need only attend to the upper and lower limits of indeterminacy, it being indifferent whether $\phi_n(x)$ has all values between these limits, or some values only. The fluctuation of $\phi_n(x)$ in any interval (α, β) is the excess of the upper limit of the numbers $\overline{\phi_n(x)}$ for all points in (α, β), over the lower limit of the numbers $\underline{\phi_n(x)}$ in the same interval. The saltus of $\phi_n(x)$ at the point x is the limit of the fluctuation in an interval $(x-\delta,\ x+\delta)$, when δ is indefinitely diminished, and this saltus is $\geqq \overline{\phi_n(x)} - \underline{\phi_n(x)}$. Riemann's theory of integration is applicable to such a function $\phi_n(x)$, when it is limited, just as in the case of a single-valued function. For any fixed value of x, the numbers $\overline{\phi_1(x)}, \overline{\phi_2(x)}, \ldots \overline{\phi_n(x)} \ldots \underline{\phi_1(x)}, \underline{\phi_2(x)}, \ldots \underline{\phi_n(x)} \ldots$ form a set which we may denote by G. Let us consider the derivative G' of G; then, if G' be limited, since it is a closed set, it has a greatest value A, and a least value B. These numbers A and B are such that, for a given ϵ, there are an infinite number of values of n such that $|\overline{\phi_n(x)} - A| < \epsilon$, and also an infinite number of values of n such that $|\underline{\phi_n(x)} - B| < \epsilon$. If G' be unlimited in one direction, or in both directions, either A, or B, or both, may be regarded as having one of the improper values ∞, $-\infty$.

We now define a function $\phi(x)$, for the interval (a, b), in the following manner:—When, for a particular value of x, the numbers A and B are equal and finite, their value is taken to be that of $\phi(x)$. If A and B are unequal and finite, we regard $\phi(x)$ as multiple-valued, with $\overline{\phi(x)} = A$, $\underline{\phi(x)} = B$. If either A or B have one of the improper values ∞, $-\infty$, the point x is taken to be a point of infinite discontinuity of $\phi(x)$. The function $\phi(x)$ is regarded as a single function, not necessarily limited, and it may have either a proper integral, or an improper integral in (a, b), in accordance with Harnack's definition of the improper integral of an unlimited function. This function $\phi(x)$ is said to be the *limiting function* defined by the sequence $\{\phi_n(x)\}$; and the functions $\phi_n(x)$ are said to *converge*, in an extended sense of the term, to the function $\phi(x)$; and thus we write $\phi(x) = \lim_{n=\infty} \phi_n(x)$.

In case the sequence $\{\overline{\phi_n(x)}\}$ be non-diminishing, so that, for every value of x and n, the condition $\overline{\phi_n(x)} \leqq \overline{\phi_{n+1}(x)}$ is satisfied, the sequence $\{\overline{\phi_n(x)}\}$ has, for each value of x, either a definite upper limit A, or else the improper limit $+\infty$. If $\underline{\phi_n(x)} \geqq \underline{\phi_{n+1}(x)}$, for every value of x and n, the sequence $\{\underline{\phi_n(x)}\}$ has, for each value of x, either a definite lower limit B, or else the improper lower limit $-\infty$.

Let a positive number ϵ, and a positive integer n_1 be arbitrarily chosen, and let E be a set of points in (a, b), of which the measure is zero. Let us suppose that, for each point x_1, in (a, b), which does not belong to a certain component E_ϵ of E, this component depending on ϵ, an integer $n\,(> n_1)$, and also a neighbourhood $(x_1 - \delta, x_1 + \delta')$ can be determined, such that the four inequalities $|\overline{\underline{\phi(x)}} - \overline{\underline{\phi_n(x)}}| < \epsilon$ are all satisfied at every point in the interval $(x_1 - \delta, x_1 + \delta')$ which is in (a, b). Then, provided this condition be satisfied for every value of ϵ; and also E be such that each point of it belongs to E_ϵ for some sufficiently small value of ϵ, the convergence of the sequence $\{\phi_n(x)\}$ to $\phi(x)$ is said to be *regular*, in the extended sense, in (a, b) *except for the set E of zero measure.*

In case, for each value of x, the sequence $\{\phi_n(x)\}$ is an increasing one, so that $\overline{\phi_n(x)} \leqq \overline{\phi_{n+1}(x)}$, and also $\underline{\phi_n(x)} \leqq \underline{\phi_{n+1}(x)}$, when the conditions $|\overline{\underline{\phi(x)}} - \overline{\underline{\phi_n(x)}}| < \epsilon$ are satisfied for a particular value of n, they are also satisfied for every greater value. In the general case however this is no longer true.

As in § 379, it is seen that the set E_ϵ must be non-dense, and therefore that the set E is of the first category.

The set E contains every point at which $\phi(x)$ has not a definite finite value, for since $\overline{\phi(x)} - \overline{\phi_n(x)}$, $\underline{\phi(x)} - \underline{\phi_n(x)}$ are both numerically less than ϵ, at a point which does not belong to E, for some value of n, it follows that $\overline{\phi(x)} - \underline{\phi(x)}$ is less than 2ϵ; and since ϵ is arbitrarily small, it follows that $\overline{\phi(x)} = \underline{\phi(x)}$. It is clear that the points of infinite discontinuity of $\phi(x)$ belong to the set E_ϵ, whatever be the value of ϵ.

As in § 380, it can be shewn that, if all the points of E_ϵ be enclosed in the interiors of intervals of a finite set, of which the sum is η, the conditions $|\overline{\underline{\phi(x)}} - \overline{\underline{\phi_n(x)}}| < \epsilon$ are satisfied at every point of (a, b) not interior to the intervals of the finite set, where n has one of a finite number of values $n_1 + p_1, n_1 + p_2, \ldots n_1 + p_r$. The particular number $n_1 + p$ which must be taken for a point x depends upon the value of x, but the same number $n_1 + p$ is applicable to all the points of one or more continuous intervals.

400. It will now be assumed that the improper double integral $\int f(x, y)\,(dx\,dy)$ exists, in accordance with the definition of § 321; and a necessary condition will be found that $\int_a^b dx \int_c^d f(x, y)\,dy$ exists.

We shall consider a sequence $f_1(x, y), f_2(x, y), \dots f_n(x, y) \dots$ of functions obtained from $f(x, y)$ as in de la Vallée-Poussin's definition of the improper double integral, given in § 321.

The integral $\int_c^d f_n(x, y)\,dy$ will be denoted by $\phi_n(x)$, where $\phi_n(x)$ may either have a determinate value, or may have as limits of indeterminacy $\overline{\phi_n(x)}$, $\underline{\phi_n(x)}$, the upper and lower values of the integral $\int_c^d f_n(x, y)\,dy$, in accordance with Darboux's definition of the upper and lower integrals of a limited function (see § 252). The existence* of $\int f_n(x, y)(dx\,dy)$ does not ensure the determinacy of $\phi_n(x)$ for all values of x. The integral

$$\int_c^d f(x, y)\,dy$$

will be denoted by $\phi(x)$; a similar remark applies to the determinacy of $\phi(x)$ as in the case of $\phi_n(x)$. Moreover $\phi(x)$ may have the improper value ∞, or $-\infty$, or may have one of these as a limit of indeterminacy; for $f(x, y)$ does not necessarily possess for each value of x either a proper or an improper integral in the interval (c, d). When, for a fixed x, the function $f(x, y)$ has points of infinite discontinuity with respect to the variable y, in the interval (c, d), the value of $\overline{\int_c^d} f(x, y)\,dy$, or $\overline{\phi(x)}$, is the upper limit of the derivative of the set of numbers $\overline{\phi_n(x)}$; also $\underline{\phi(x)}$ is the lower limit of the derivative of the set of numbers $\underline{\phi_n(x)}$. It may happen that $\overline{\phi(x)}$, or $\underline{\phi(x)}$, has an improper value ∞ or $-\infty$. In the present case all the functions $\phi_n(x)$ are limited functions.

Since $f(x, y)$ is integrable in the fundamental rectangle, all the functions $f_n(x, y)$ have proper integrals in that domain. The proper integral

$$\int f_n(x, y)(dx\,dy),$$

is, by the theorem of § 314, replaceable by the repeated integral

$$\int_a^b dx \int_c^d f_n(x, y)\,dy,$$

and thus $\phi_n(x)$ is integrable in the linear interval (a, b). It follows that the points of discontinuity of $\phi_n(x)$ form a set of points of linear measure zero. The set of all points of discontinuity of any of the functions

$$\phi_1(x),\ \phi_2(x), \dots \phi_n(x) \dots,$$

* In de la Vallée-Poussin's investigation in *Liouville's Journal*, ser. 4, vol. VIII, the restrictive assumptions are made that $\phi_n(x)$ and $\phi(x)$ are everywhere definite and finite.

is consequently also a set of zero measure. If $\phi(x)$ be integrable in the interval (a, b), its points of discontinuity must form a set of zero measure.

Let us suppose that $\phi(x)$ is integrable in (a, b), and thus that

$$\int_a^b dx \int_c^d f(x, y)\, dy$$

exists; and let us assume, that, if possible, the set E_ϵ, referred to in the definition of regular convergence, has its measure greater than zero. Remove from E_ϵ those points at which one or more of the functions

$$\phi_1(x),\quad \phi_2(x), \ldots \phi_n(x) \ldots$$

is discontinuous, and also remove all those points at which $\phi(x)$ is discontinuous. We have then left a set F_ϵ, of measure equal to that of E_ϵ, and therefore by hypothesis greater than zero. At every point of F_ϵ all the functions $\phi_n(x)$ are definite and continuous, and $\phi(x)$ is also definite and continuous. If ξ be a point of F_ϵ, the number $n (> n_1)$ can be so chosen that

$$|\phi(\xi) - \phi_n(\xi)| < \tfrac{1}{3}\epsilon;$$

also δ can be so chosen that, for every x in the interval $(\xi - \delta, \xi + \delta)$, the four inequalities

$$|\phi(\xi) - \overline{\underline{\phi(x)}}| < \tfrac{1}{3}\epsilon,$$

$$|\phi_n(\xi) - \overline{\underline{\phi_n(x)}}| < \tfrac{1}{3}\epsilon$$

are all satisfied. From these inequalities we deduce that the four inequalities $|\overline{\underline{\phi(x)}} - \overline{\underline{\phi_n(x)}}| < \epsilon$ are all satisfied for all points x in the interval

$$(\xi - \delta,\ \xi + \delta).$$

But this is contrary to the hypothesis that ξ is a point belonging to E_ϵ. It therefore follows that, on the assumption that $f(x, y)$ has an improper integral in the fundamental rectangle, the repeated integral

$$\int_a^b dx \int_c^d f(x, y)\, dy$$

cannot exist unless E_ϵ has the measure zero. Since this holds for every ϵ, we have obtained the following theorem:—

If $f(x, y)$ have an improper (absolutely convergent) integral in the fundamental rectangle, a necessary condition for the existence of the repeated integral $\int_a^b dx \int_c^d f(x, y)\, dy$ is that the convergence of

$$\int_c^d f_n(x, y)\, dy \quad \text{to} \quad \int_c^d f(x, y)\, dy$$

*should be regular, except for a set of points E, of the first category, and of zero measure**.

401. Let it now be assumed that $f(x, y) \geqq 0$, throughout the fundamental rectangle. It will be shewn that, in this case, the condition of regular convergence of $\{\phi_n(x)\}$ to $\phi(x)$, at all points except a set of the first category, and of zero measure, is sufficient to ensure that $\int_a^b dx \int_c^d f(x, y)\, dy$ exists, and that it is equal to $\int f(x, y)(dx\, dy)$; it being assumed that the double integral exists.

In this case the four inequalities $|\overline{\phi(x)} - \overline{\phi_n(x)}| < \epsilon$, are equivalent to the one $\overline{\phi(x)} - \underline{\phi_n(x)} < \epsilon$; also if, at any point x, this is satisfied for a value of n, then it is also satisfied for all greater values of n. Including all the points of E_ϵ in the interior of intervals of a finite set, such that the sum of these intervals is the arbitrarily small number η, we see that the condition $\overline{\phi(x)} - \underline{\phi_n(x)} < \epsilon$ is satisfied for one and the same value of $n\,(> n_1)$ at all points x not interior to the intervals whose sum is η. For we have only to take for n the greatest of the numbers $n_1 + p_1,\ n_1 + p_2, \dots n_1 + p_r$ defined in § 399. The number ϵ being fixed, we can choose η so small that the double integral $\int f(x, y)(dx\, dy)$ over those rectangles of which the height is $d - c$, and the sum of the breadths η, is less than an arbitrarily fixed positive number ζ; this follows from de la Vallée-Poussin's definition of an improper integral. For m may be chosen so great that

$$\int f(x, y)(dx\, dy) - \int f_m(x, y)(dx dy) < \tfrac{1}{2}\zeta,$$

when the integrals are taken over the fundamental rectangle, and therefore also when they are both taken over any prescribed part of that rectangle; now η can be so chosen that $\int f_m(x, y)(dx\, dy) < \frac{1}{2}\zeta$, and therefore so that $\int f(x, y)(dx\, dy) < \zeta$, the integrals being taken over the rectangles of total breadth η. The number η being thus fixed, a number m exists, such that for $n \geqq m$, we have $\overline{\phi(x)} - \underline{\phi_n(x)} < \epsilon$, except in the interiors of the intervals which enclose E_ϵ. We have therefore

$$\overline{\int}\phi(x)\, dx - \underline{\int}\phi_n(x)\, dx < \epsilon(b - a - \eta) < \epsilon(b - a),$$

* This theorem was established by de la Vallée-Poussin only for the case in which $f(x, y) \geqq 0$, and only under a restrictive hypothesis, that $\int_c^d f(x, y)\, dy$ and $\int_c^d f_n(x, y)\, dy$ both have definite finite values at all points x not belonging to a set of points of zero *content*; whereas this set need only have zero measure.

the integration being taken along the parts of (a, b) which remain when the enclosing intervals are removed. Hence we have

$$\overline{\int}\phi(x)\,dx - \int f_n(x, y)\,(dx\,dy) < \epsilon(b-a),$$

where the double integral is taken over the fundamental rectangle with the exception of those rectangles of which the breadths are the enclosing intervals of sum η. Also, if ζ' be an arbitrarily chosen positive number, we can choose n so great that

$$\int f(x, y)\,(dx\,dy) - \int f_n(x, y)\,(dx\,dy) < \zeta',$$

where both double integrals are taken over the same region as before. We now see that

$$\left|\overline{\int}\phi(x)\,dx - \int f(x, y)\,(dx\,dy)\right| < \zeta' + \epsilon(b-a),$$

and hence we have

$$\left|\int f(x, y)\,(dx\,dy) - \overline{\int}\phi(x)\,dx\right| < \zeta + \zeta' + \epsilon(b-a),$$

where the double integral is now taken over the whole fundamental rectangle, and the single integral over (a, b) with the exception of those parts whose sum is η. Now ζ' is arbitrarily small, and ζ, η converge together to zero. It follows that $\int_a^b \phi(x)\,dx$, whether definite or not, lies between

$$\int f(x, y)\,(dx\,dy) \pm \epsilon(b-a);$$

and since ϵ is arbitrarily small, it follows that $\int_a^b \phi(x)\,dx$ exists as a definite proper or improper integral, and is equal to

$$\int f(x, y)\,(dx\,dy).$$

The following theorem has now been established:—

If $f(x, y) \geqq 0$, and the function have an improper double integral in the fundamental rectangle, then the condition that $\int_c^d f_n(x, y)\,dy$ converges regularly to $\int_c^d f(x, y)\,dy$, except for a set of points E, of the first category, and of zero measure, is a sufficient condition that $\int_a^b dx \int_c^d f(x, y)\,dy$ exists, and is equal to $\int f(x, y)\,(dx\,dy)$.

The sufficiency of the same condition, for the case in which $f(x, y)$ is not restricted to have one sign only, does not appear to be capable of establishment, because it is in this case impossible to shew that the conditions

$$|\overline{\phi(x)} - \overline{\phi_n(x)}| < \epsilon,$$

are satisfied at all points except in the enclosing intervals, for one and the same value of n; it having been only established that they hold when n has one of a finite number of values.

Combining the present result with that of § 400, we see that:—

If $f(x, y) \geqq 0$, and have an absolutely convergent improper integral in the fundamental rectangle, the necessary and sufficient condition that

$$\int_a^b dx \int_c^d f(x, y)\,dy$$

should exist, and be equal to $\int f(x, y)\,(dx\,dy)$, is that $\int_c^d f_n(x, y)\,dy$ should converge regularly to $\int_c^d f(x, y)\,dy$, except for a set E of the first category and of zero measure.

It has also been established that, *when $\int f(x, y)\,(dx\,dy)$ exists, then if $\int_a^b dx \int_c^d f(x, y)\,dy$ have a definite meaning, it is equal to the double integral.*

For it has been shewn in § 400, that the repeated integral cannot have a definite meaning, $\phi(x)$ being integrable in (a, b), unless the convergence is of the kind specified; and when $f(x, y) \geqq 0$, this is sufficient that the repeated integral may be equal to the double integral.

402. Returning to the case in which $f(x, y)$ is not restricted to be of one sign, the following theorem will be established:—

If $f(x, y)$ have an absolutely convergent improper integral in the fundamental rectangle, a sufficient condition that $\int_a^b dx \int_c^d f(x, y)\,dy$ may exist, and may have the same value as the double integral $\int f(x, y)\,(dx\,dy)$, is that

$$\int_c^d |f_n(x, y)|\,dy$$

shall converge regularly to $\int_c^d |f(x, y)|\,dy$, except for a set of points of the first category and of zero measure.

Employing $f(x, y) = f^+(x, y) - f^-(x, y)$, $f_n(x, y) = f_n^+(x, y) - f_n^-(x, y)$, and denoting $\int_c^d f_n^+(x, y)\,dy$, $\int_c^d f_n^-(x, y)\,dy$ by $\phi_n^+(x)$, $\phi_n^-(x)$ respectively, we see that the condition stated in the theorem is that $\phi_n^+(x) + \phi_n^-(x)$ converges regularly to $\phi^+(x) + \phi^-(x)$. In order that this condition may be satisfied we must have $\overline{\phi^+(x)} + \overline{\phi^-(x)} - \underline{\phi_n}^+(x) - \underline{\phi_n}^-(x) < \epsilon$, for a sufficiently great value of n, at every point of (a, b) not interior to a finite set of intervals of arbitrarily small sum η enclosing the points of E_ϵ, a set of zero content. From this condition we deduce that

$$\overline{\phi^+(x)} - \overline{\phi_n^+(x)} < \epsilon, \text{ and } \overline{\phi^-}(x) - \overline{\phi_n}^-(x) < \epsilon,$$

at every point not in the interior of the intervals, for all sufficiently great values of n.

Hence it follows that $\phi_n^+(x)$ converges regularly to $\phi^+(x)$, and also $\phi_n^-(x)$ converges regularly to $\phi^-(x)$, at all points except those of a set E, of zero measure. It follows from the theorem of § 401, that the two repeated integrals

$$\int_a^b dx \int_c^d f^+(x, y)\, dy, \quad \int_a^b dx \int_c^d f^-(x, y)\, dy$$

exist, and are equal to

$$\int f^+(x, y)\,(dx\, dy), \quad \int f^-(x, y)\,(dx\, dy)$$

respectively; and from this it follows that

$$\int_a^b dx \int_c^d f(x, y)\, dy \text{ exists, and } = \int f(x, y)\,(dx\, dy).$$

The condition stated in the theorem, though sufficient, is not necessary; for the integral $\int_a^b \{\phi^+(x) - \phi^-(x)\}\, dx$ may exist only as a non-absolutely convergent improper integral, in which case

$$\int_a^b \{\phi^+(x) + \phi^-(x)\}\, dx$$

does not exist. In this case

$$\int_a^b dx \int_c^d |f(x, y)|\, dy$$

not being existent, the convergence of

$$\int_c^d |f_n(x, y)|\, dy \text{ to } \int_c^d |f(x, y)|\, dy$$

cannot be regular.

403. Whether the double integral $\int f(x, y)\,(dx\, dy)$ exist or not, the proof of the theorem in § 400, suffices to shew that, if all the double integrals

$\int f_n(x, y)(dx dy)$ exist, then it is a necessary condition for the existence of the repeated integral $\int_a^b dx \int_c^d f(x, y)\, dy$, that the integrals

$$\int_c^d f_n(x, y)\, dy$$

should converge regularly to $\int_c^d f(x, y)\, dy$, except for a set of points of the first category and of zero measure.

Moreover, if it be known that $\int_c^d f(x, y)\, dy$ is a function of x, which is limited in the interval (a, b), we can infer the existence of the double integral $\int f(x, y)(dx\, dy)$. For since

$$\int f_n(x, y)(dx\, dy) = \int_a^b dx \int_c^d f_n(x, y)\, dy,$$

we have

$$\left| \int f_n(x, y)(dx dy) \right| < (b-a)\, U,$$

where U is the upper limit of $\left| \int_c^d f_n(x, y)\, dy \right|$ in the interval (a, b). It is thus seen that $\int f_n(x, y)(dx\, dy)$ cannot increase indefinitely in numerical value, as n is increased indefinitely. The following theorem has thus been established:—

If all the functions $f_n(x, y)$ have double integrals in the fundamental rectangle, and $\int_c^d f_n(x, y)\, dy$ converges to $\int_c^d f(x, y)\, dy$ regularly for all values of x, except for a set of zero measure and of the first category, then, if $\int_c^d f(x, y)\, dy$ be a limited function of x in the interval (a, b), the double integral $\int f(x, y)(dx dy)$ exists, and is equal to

$$\int_a^b dx \int_c^d f(x, y)\, dy.$$

Combining this theorem with that of § 402, we obtain the following theorem:—

If all the functions $f_n(x, y)$ have double integrals in the fundamental rectangle, and either $\int_c^d f(x, y)\, dy$ is limited in the interval (a, b) of x, or

$\int_a^b f(x, y)\, dx$ *is limited in the interval* (c, d) *of* y, *then if the conditions are satisfied, that the sequences*

$$\int_c^d |f_n(x, y)|\, dy, \quad \int_a^b |f_n(x, y)|\, dx$$

converge to the limits

$$\int_c^d |f(x, y)|\, dy, \quad \int_a^b |f(x, y)|\, dx,$$

in each case regularly, except for a set of points of zero measure and of the first category, then the double integral exists, and

$$\int f(x, y)\,(dx\, dy) = \int_a^b dx \int_c^d f(x, y)\, dy = \int_c^d dy \int_a^b f(x, y)\, dx.$$

Further investigations have been given by de la Vallée-Poussin of sufficient conditions for the equality of the two repeated integrals, without it being assumed that the double integral exists. In these investigations somewhat complicated restrictions are made as to the mode of distribution of the points of infinite discontinuity of the function. Moreover other assumptions are implicitly made as to the definiteness of the single integral $\int_c^d f(x, y)\, dy$ for all values of x, and of $\int_a^b f(x, y)\, dx$ for all values of y.

REPEATED LEBESGUE INTEGRALS.

404. The subject of repeated integration will now be considered for the case of a function which possesses a Lebesgue double integral. Some further definitions will be first required.

The function $f(x, y)$ being a summable function, defined as in § 287, for the points of the rectangle bounded by $x = a$, $x = b$, $y = c$, $y = d$, it is to be observed that we have no assurance that the function $f(x_0, y)$, defined for all values of y on the straight line $x = x_0$, is a summable function of y; unless indeed we assume that all functions are summable. For it is not certain that the section $E(x_0)$, of a measurable two-dimensional set of points E, by the straight line $x = x_0$, is linearly measurable. It will be convenient to denote the linear measure of a set e of points on a straight line by $m_l(e)$, in order to distinguish linear measure from the plane measure $m(E)$ of a set in the plane.

Let $F(x)$ be a limited function defined for the points of a measurable set e of points in the x-axis. If $F(x)$ be a summable function, then the Lebesgue integral of $F(x)$ taken over the set e exists, in accordance with the definition given in § 287. For the case in which $F(x)$ is not summable, Lebesgue has defined upper and lower integrals in e, which are however quite distinct conceptions from the upper and lower integrals as defined in § 252, in connection with integration as defined by Riemann. For any summable function, defined for the interval (a, b), the upper and lower integrals defined by Lebesgue and denoted by

$$\int_{a\,\text{sup}}^{b} F(x)\,dx, \quad \int_{a\,\text{inf}}^{b} F(x)\,dx,$$

have one and the same value, whereas the upper and lower integrals

$$\overline{\int_a^b} F(x)\,dx, \quad \underline{\int_a^b} F(x)\,dx,$$

in accordance with the definition of § 252, have different values, unless $F(x)$ possesses a Riemann integral in the interval (a, b).

If $\phi(x)$ denote any limited summable function defined for the set e, for which $F(x)$ is defined, and such that $\phi(x) > F(x)$ for every point of e, then the lower limit of the Lebesgue integral $\int_e \phi(x)\,dx$, for all possible functions $\phi(x)$ which satisfy the prescribed condition, is defined to be the upper Lebesgue integral of $F(x)$ taken over e, and is denoted by

$$\int_{e\,\text{sup}} F(x)\,dx.$$

The lower Lebesgue integral $\int_{\text{inf}} F(x)\,dx$ is the upper limit of the integrals of all limited summable functions $\phi(x)$, such that $F(x) > \phi(x)$.

It has been shewn in § 83, that any linear measurable set e contains a set e_1 of the same measure as itself, and such that e_1 consists of all those points which are common to all sets of intervals belonging to a sequence of such sets. It has also been shewn that e is contained in a set e_2, of measure equal to $m_l(e)$, of the same type as the set e_1.

Exactly similar reasoning to that in § 83 suffices to shew that a plane set E contains a set E_1, such that $m(E) = m(E_1)$; where E_1 is that set of points which is common to all sets of rectangles belonging to a sequence of such sets of rectangles. The successive sets of the sequence may be taken each to contain the next; and in each set the rectangles do not overlap. The set E is also contained in a set* E_2, such that $m(E_2) = m(E)$, and such that E_2 is a set of the same type as E_1.

405. Let the function $\phi(x, y)$ be defined to be $= 1$, for all points of the measurable set E, and to be zero at all other points of the fundamental rectangle. If E_1, E_2 be the sets of measures equal to that of E, E_1 being

* The sets e_1, e_2, E_1, E_2 are of the kind described by Lebesgue as "measurable (B)."

contained in E, and E_2 containing E, as explained in § 404; then the sections $E_1(x_0)$, $E_2(x_0)$ of E_1 and E_2, by the straight line $x = x_0$, are both measurable. It is clear that

$$m_l[E_1(x_0)] \leqq \underset{\text{inf}}{\int_c^d} \phi(x_0, y)\, dy,$$

$$m_l[E_2(x_0)] \geqq \underset{\text{sup}}{\int_c^d} \phi(x_0, y)\, dy,$$

the integrals being taken over the interval (c, d) of y.

If $D_1, D_2, \ldots D_\iota, \ldots$ be a sequence of sets of rectangles, such that each set contains the next, and if $D_{\iota j}$ denote a single rectangle of the set D_ι, we may denote the section of $D_{\iota j}$ by the straight line $x = x_0$, by $D_{\iota j}(x_0)$; also we may denote by $D_\iota(x_0)$ the section of the set D_ι. We have then

$$m_l[D_\iota(x_0)] = \sum_{j=1}^{\infty} m_l[D_{\iota j}(x_0)].$$

Since $\sum_{j=1}^{j} m_l[D_{\iota j}(x_0)]$ is limited, for all values of j and x_0, we may integrate term by term; hence we have

$$m(D_\iota) = \sum_{j=1}^{\infty} \int_a^b m_l[D_{\iota j}(x)]\, dx$$

$$= \int_a^b m_l[D_\iota(x)]\, dx,$$

in virtue of the theorem of § 384.

Again, we have

$$m(E) = m(E_1) = \lim_{\iota=\infty} m(D_\iota)$$

$$= \lim_{\iota=\infty} \int_a^b m_l[D_\iota(x)]\, dx;$$

and since $m_l[D_\iota(x)]$ is limited, for all values of ι and x, we have

$$m(E) = \int_a^b m_l[E_1(x)]\, dx,$$

if we take the sequence $\{D_\iota\}$ to be the one which defines E_1.

It follows that

$$m(E) \leqq \underset{\text{inf}}{\int_a^b} m_{l,\,\text{int}}[E(x)]\, dx,$$

where $m_{l,\,\text{int}}$ denotes the interior measure of the set; and therefore

$$\int_A \phi(x, y)(dx\,dy) \leqq \underset{\text{inf}}{\int_a^b} dx \underset{\text{inf}}{\int_c^d} \phi(x, y)\, dy.$$

In a similar manner, it can be proved that

$$\int_A \phi(x, y)(dx\,dy) \geqq \underset{\text{sup}}{\int_a^b} dx \underset{\text{sup}}{\int_c^d} \phi(x, y)\,dy.$$

It follows from the two relations, that

$$\int_A \phi(x, y)(dxdy) = \underset{\text{sup}}{\int_a^b} dx \underset{\text{sup}}{\int_c^d} \phi(x, y)\,dy = \underset{\text{inf}}{\int_a^b} dx \underset{\text{inf}}{\int_c^d} \phi(x, y)\,dy,$$

where $\phi(x, y) = 1$, at the points of E, and $= 0$, at the points of $C(E)$.

Denoting by U and L, the upper and lower limits of $f(x, y)$ in the rectangle, let (L, U) be divided into parts $(a_0, a_1), (a_1, a_2), \ldots (a_{n-1}, a_n)$, where $a_0 = L$, $a_n = U$; and let α denote the greatest of the differences $a_p - a_{p-1}$, for $p = 1, 2, 3, \ldots n$. Let E_p be the set of points in the rectangle at which $f(x, y) = a_p$, and let E_p' denote that set for which $a_p < f(x, y) < a_{p+1}$. Also let $f_p(x, y)$ denote a function which is equal to $f(x, y)$ at all points of the set E_p, and is zero at all other points; and let $f_p'(x, y)$ denote a function which is equal to $f(x, y)$ at points of the set E_p', and is zero at all other points.

We have then

$$\int f(x, y)(dxdy) = \sum_{p=0}^{n} \int f_p(x, y)(dxdy) + \sum_{p=0}^{n-1} \int f_p'(x, y)(dxdy).$$

Now $$\int f_p(x, y)(dxdy) = a_p m(E_p) = \underset{\text{inf}}{\int_a^b} dx \underset{\text{inf}}{\int_c^d} f_p(x, y)\,dy,$$

by the theorem proved above.

Also $\int f_p'(x, y)(dxdy)$ is between $a_p m(E_p')$ and $a_{p+1} m(E_p')$, or between

$$\underset{\text{inf}}{\int_a^b} dx \underset{\text{inf}}{\int_c^d} \phi(x, y)\,dy \text{ and } \underset{\text{inf}}{\int_a^b} dx \underset{\text{inf}}{\int_c^d} \psi(x, y)\,dy;$$

where $\phi(x, y) = a_p$, $\psi(x, y) = a_{p+1}$ at the points of E_p', and where both functions vanish at all other points. These two repeated integrals differ from one another by less than $\alpha m(E_p')$; also $\underset{\text{inf}}{\int_a^b} dx \underset{\text{inf}}{\int_c^d} f_p'(x, y)\,dy$ differs from either by less than $\alpha m(E_p')$. Hence we have

$$\left| \int f_p'(x, y)(dxdy) - \underset{\text{inf}}{\int_a^b} dx \underset{\text{inf}}{\int_c^d} f_p'(x, y)\,dy \right| < \alpha m(E_p').$$

It follows that

$$\int_A f(x, y)(dxdy) - \sum_{p=0}^{n} \underset{\text{inf}}{\int_a^b} dx \underset{\text{inf}}{\int_c^d} f_p(x, y)\,dy - \sum_{p=0}^{n-1} \underset{\text{inf}}{\int_a^b} dx \underset{\text{inf}}{\int_c^d} f_p'(x, y)\,dy,$$

is numerically less than αA. We next see that

$$\underset{\text{inf}}{\int_c^d} f(x, y)\,dy \geqq \sum_{p=0}^{n} \underset{\text{inf}}{\int_c^d} f_p(x, y)\,dy + \sum_{p=0}^{n-1} \underset{\text{inf}}{\int_c^d} f_p'(x, y)\,dy.$$

For, if possible, let

$$\underset{\inf}{\int_c^d} f(x, y)\, dy = \sum_{p=0}^{n} \underset{\inf}{\int_c^d} f_p(x, y)\, dy + \sum_{p=0}^{n-1} \underset{\inf}{\int_c^d} f_p'(x, y)\, dy - \zeta,$$

where ζ is some positive number; then summable functions

$$\bar{f}_p(x, y), \quad \bar{f}_p'(x, y)$$

such that

$$f_p(x, y) > \bar{f}_p(x, y), \qquad f_p'(x, y) > \bar{f}_p'(x, y),$$

can be so determined that

$$\underset{\inf}{\int_c^d} f_p(x, y)\, dy - \int_c^d \bar{f}_p(x, y)\, dy < \zeta/(2n+1)$$

with a similar inequality for the case of the functions

$$f_p'(x, y), \quad \bar{f}_p'(x, y);$$

we have then

$$\underset{\inf}{\int_c^d} f(x, y)\, dy < \int_c^d \left\{ \sum_{p=0}^{n} \bar{f}_p(x, y) + \sum_{p=0}^{n-1} \bar{f}_p'(x, y) \right\} dy,$$

whereas $f(x, y)$ is greater than the summable integrand in the integral on the right-hand side. This is contrary to the definition of the lower integral, and therefore the positive number ζ cannot exist. A repetition of the same reasoning suffices to shew* that

$$\underset{\inf}{\int_a^b} dx \underset{\inf}{\int_c^d} f(x, y)\, dy \geqq \sum_{p=0}^{n} \underset{\inf}{\int_a^b} dx \underset{\inf}{\int_c^d} f_p(x, y)\, dy + \sum_{p=0}^{n-1} \underset{\inf}{\int_a^b} dx \int_c^d f_p'(x, y)\, dy.$$

We see now that

$$\int_A f(x, y)(dx\, dy) < \underset{\inf}{\int_a^b} dx \underset{\inf}{\int_c^d} f(x, y)\, dy + \alpha A;$$

and as α is arbitrarily small, this shews that

$$\int_A f(x, y)(dx dy) \leqq \underset{\inf}{\int_a^b} dx \underset{\inf}{\int_c^d} f(x, y)\, dy.$$

It may, in a similar manner, be proved that

$$\int_A f(x, y)(dx dy) \geqq \underset{\sup}{\int_a^b} dx \underset{\sup}{\int_c^d} f(x, y)\, dy.$$

From these relations we deduce the theorem†

$$\int_A f(x, y)(dx dy) = \underset{\sup}{\int_a^b} dx \underset{\sup}{\int_c^d} f(x, y)\, dy = \underset{\inf}{\int_a^b} dx \underset{\inf}{\int_c^d} f(x, y)\, dy,$$

where $f(x, y)$ is any limited function, summable in the plane.

* Lebesgue's statement (*loc. cit.*, p. 279) is the reverse of the inequality here given. It would appear that this is due to accident, as his result does not follow from the inequality he gives.

† Lebesgue, *Annali di Mat.* ser. 3, vol. VII, p. 278.

406. From the theorem just established, and considering the corresponding repeated integrals taken first with respect to x and then with respect to y, we have the following theorem :—

If $f(x, y)$ be any limited summable function, defined in the rectangle A, then

$$\int_A f(x, y)(dxdy) = \int_a^b dx \int_c^d f(x, y)\,dy = \int_c^d dy \int_a^b f(x, y)\,dx,$$

whenever the repeated integrals have definite meanings, in accordance with Lebesgue's definition of integration. It may happen that only one of the repeated integrals has a meaning.

In some cases the function $f(x, y)$ may be such that both the repeated integrals exist in accordance with Riemann's definition of integration. In case the function have a double integral in accordance with the extended Riemann definition for a function of two variables, then the theorem reduces to the one established in § 314. It may however happen that $f(x, y)$ has a double integral only in accordance with Lebesgue's definition. This throws light upon the questions considered in § 316. We may, in fact, state the following theorems :—

If the repeated integrals of a limited summable function $f(x, y)$ both exist, in accordance with Riemann's definition, then they are equal to one another, and to the Lebesgue double integral of $f(x, y)$.

If only one of the repeated integrals of a limited summable function exist, in accordance with Riemann's definition, then the other may exist in accordance with Lebesgue's definition, and they are then equal to one another, and to the Lebesgue double integral of the function.

The only possible case in which both repeated integrals can exist and have unequal values is when the function is not summable in the plane.

EXAMPLES.

1. For the function defined in § 316, Ex. 1, only one of the repeated integrals exists, in accordance with the definition there employed; neither does the double integral exist. The Lebesgue double integral exists, and $=1$. For the set of points at which $f(x, y)=1$ has the measure zero; and therefore the function has the same Lebesgue integral as that function which, at every point (x, y), has the value $2y$. The functional values at a set of points of zero measure are irrelevant in a Lebesgue integral. The other repeated integral also exists, in accordance with the definition of Lebesgue, as may be easily verified.

2. For the function defined in § 316, Ex. 4, both the repeated integrals exist, in accordance with the definition there employed, and they have the value c; the double integral, however, does not exist. But the Lebesgue double integral exists, and $=c$; for the points at which $f(x, y)=c'$, although they are everywhere-dense, form a set of plane measure zero.

407. When $f(x, y)$ is unlimited in the rectangle for which it is defined, U may have the improper value ∞, and L may have the improper value $-\infty$. No essential change is required in the reasoning of § 405, which suffices to shew that

$$\int_A f(x, y)\,(dx\,dy) = \int_{a\atop\text{inf}}^{b} dx \int_{c\atop\text{inf}}^{d} f(x, y)\,dy = \int_{a\atop\text{sup}}^{b} dx \int_{c\atop\text{sup}}^{d} f(x, y)\,dy,$$

whenever the integrals have a meaning. The result

$$\int_A f(x, y)\,(dx\,dy) = \int_a^b dx \int_c^d f(x, y)\,dy = \int_c^d dy \int_a^b f(x, y)\,dx,$$

also holds whenever the integrals have a meaning.

The integration of $\int_a^b f(x, y)\,dx$ with respect to y between limits (c, y_1) is frequently performed by reversing the order of the integrations; thus the value of $\int_c^{y_1} dy \int_a^b f(x, y)\,dx$ is frequently calculated by means of

$$\int_a^b dx \int_c^{y_1} f(x, y)\,dy.$$

This process is frequently spoken of as "integrating under the integral sign with respect to a parameter y." What has been just established enables us to state the following theorem:—

Integration with respect to a parameter, under the integral sign, is always valid, provided the function is summable and integrable in the plane, if the result of the process have a definite meaning.

It must, however, be remembered that even if $\int_c^{y_1} dy \int_a^b f(x, y)\,dx$ exist, in accordance with Riemann's definition or one of its extensions,

$$\int_a^b dx \int_c^{y_1} f(x, y)\,dy$$

may exist, if it exists at all, only when Lebesgue's definition is employed.

The only case in which the repeated integrals of an unlimited function can both exist, but have unequal values, is when the function is either not summable in the plane, or is summable and still does not possess a Lebesgue integral.

REPEATED INTEGRALS OF UNLIMITED FUNCTIONS.

408. In order to find sufficient conditions for the existence and equality of the repeated integrals of an unlimited function, which shall not depend upon the necessary existence of the double integral, in accordance with either the definitions of Jordan and de la Vallée-Poussin, or with that of Lebesgue, we recur to the wider definition of a double integral which has been given in

§ 320. That definition differs from Jordan's definition in § 318, in that rectangles only are employed for the purpose of enclosing the points of infinite discontinuity; and thus the domains D_n in Jordan's definition are to be restricted to consist each of a finite set of rectangles with sides parallel to the axes. An improper integral which exists in accordance with the definition of § 320, we shall speak of as *a restricted Jordan double integral.*

Assuming that the integral $\int_A f(x, y)(dx\,dy)$ exists, as a restricted Jordan double integral, let $f_n(x, y)$ be that limited function which, in the domain D_n, consisting of a finite set of rectangles, is equal to $f(x, y)$, and is zero in $C(D_n)$, which contains all the points of infinite discontinuity of $f(x, y)$.

We have then

$$\int_A f(x, y)(dxdy) = \lim_{n=\infty} \int_A f_n(x, y)(dxdy)$$
$$= \lim_{n=\infty} \int_a^b dx \int_c^d f_n(x, y)\,dy,$$

and therefore we have

$$\int_A f(x, y)(dxdy) = \int_a^b dx \int_c^d f(x, y)\,dy,$$

provided $$\lim_{n=\infty} \int_a^b dx \int_{\Lambda_n(x)} f(x, y)\,dy = 0,$$

where $\Lambda_n(x)$ denotes that finite set of intervals which forms the section of $C(D_n)$ by the ordinate corresponding to the abscissa x.

From this result, the following theorem, very similar to one given by Jordan*, and specifying a particular mode of satisfying the last condition, may be deduced:—

For the existence and equality of the two repeated integrals

$$\int_a^b dx \int_c^d f(x, y)\,dy, \qquad \int_c^d dy \int_a^b f(x, y)\,dx,$$

it is sufficient

(1) *That the function $f(x, y)$ possess a restricted Jordan double integral in the fundamental rectangle.*

(2) *That the points of infinite discontinuity of $f(x, y)$ be distributed on a limited number of arcs of continuous curves representing monotone functions.*

(3) *That, corresponding to any fixed positive number ϵ, positive numbers h_1, k_1 exist, such that*

$$\left|\int_x^{x+h} f(x, y)\,dx\right| < \epsilon, \qquad \left|\int_y^{y+k} f(x, y)\,dy\right| < \epsilon,$$

* See *Cours d'Analyse*, vol. II, p. 67.

for $|h| < h_1$, $|k| < k_1$, *and for every value of* x *and* y *in the fundamental rectangle.*

To shew that, under the conditions stated,

$$\lim_{n=\infty} \int_a^b dx \int_{\Lambda_n(x)} f(x, y)\, dy = 0,$$

it is clear that the points of any one such curve can be enclosed in the interiors of a finite number of rectangles, the height of each of which is $< k_1$. Then $\Lambda_n(x)$ consists of a number of intervals not exceeding the number r of the curves on which the points of infinite discontinuity lie.

We have then $\left|\int_{\Lambda_n(x)} f(x, y)\, dy\right| < r\epsilon$; and therefore

$$\left|\int_a^b dx \int_{\Lambda_n(x)} f(x, y)\, dy\right|$$

is less than the arbitrarily small number $r\epsilon(b-a)$. Thus

$$\int_a^b dx \int_c^d f(x, y)\, dy$$

exists, and is equal to the restricted Jordan integral. Similarly it can be shewn that the other repeated integral has the same value.

EXAMPLES.

1. Let $f(x, y) = \dfrac{x^2-y^2}{(x^2+y^2)^2}$, and let the domain be bounded by $x=0$, $x=1$, $y=0$, $y=1$. Neither the double integral, nor the restricted Jordan integral, exists in this case. For, if the rectangle bounded by $x=0$, $x=a$, $y=0$, $y=\beta$ be excluded from the domain, the double integral over the remainder is equal to

$$\int_0^1 dx \int_\beta^1 \frac{x^2-y^2}{(x^2+y^2)^2}\, dy + \int_a^1 dx \int_0^\beta \frac{x^2-y^2}{(x^2+y^2)^2}\, dy,$$

or to

$$\int_0^1 \left(\frac{1}{x^2+1} - \frac{\beta}{x^2+\beta^2}\right) dx + \int_a^1 \frac{\beta}{x^2+\beta^2}\, dx,$$

which is $\frac{1}{4}\pi - \tan^{-1}\dfrac{a}{\beta}$; and this has no definite limit as a, β converge independently to zero.

The repeated integrals exist, and have unequal values; for

$$\int_0^1 dx \int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2}\, dy = \int_0^1 \frac{1}{x^2+1}\, dx = \tfrac{1}{4}\pi,$$

$$\int_0^1 dy \int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2}\, dx = -\int_0^1 \frac{1}{y^2+1}\, dy = -\tfrac{1}{4}\pi.$$

2. It has been shewn in the Example, § 320, that the double integral of $\dfrac{1}{x}\sin\dfrac{1}{x}$ over

the domain bounded by $x=0$, $x=a$, $y=0$, $y=b$, does not exist. In this case however the restricted Jordan double integral exists. For

$$\int_0^b dy \int_\epsilon^a \frac{1}{x} \sin \frac{1}{x}\, dx = b \int_\epsilon^a \frac{1}{x} \sin \frac{1}{x}\, dx,$$

and this has a definite limit, for $\epsilon = 0$.

The two repeated integrals exist, and are equal to the restricted double integral. The condition

$$\lim \int_0^b dy \int_{\Lambda_n(x)} \frac{1}{x} \sin \frac{1}{x}\, dx,$$

of § 408, is satisfied, for $\Lambda_n(x)$ is in this case independent of x, and consists of an interval $(0, \epsilon)$.

3. Let $f(x, y) = (x-y)^{-\frac{2}{3}}$, and let the domain be bounded by $x=a$, $x=b$, $y=0$, $y=c$, where $c>a$. In this case the double integral exists.

For $\int_0^c (x-y)^{-\frac{2}{3}}\, dy = 3x^{\frac{1}{3}} + 3(c-x)^{\frac{1}{3}} - 3 \lim\limits_{\epsilon=0} \epsilon^{\frac{1}{3}} - 3 \lim\limits_{\epsilon'=0} \epsilon'^{\frac{1}{3}}$, and this is $3x^{\frac{1}{3}} + 3(c-x)^{\frac{1}{3}}$;

therefore $\left| \int_0^c (x-y)^{-\frac{2}{3}}\, dy \right|$ has a finite upper limit for all values of x in (a, b). It then follows from the theorem of § 403, that the double integral exists. Also since

$$\int_a^b dx \int_0^c (x-y)^{-\frac{2}{3}} dy$$

has a definite meaning, its value is the same as that of the double integral. Clearly the other repeated integral exists, and has the same value as the double integral.

4. Let the function* $\psi(x)$ be defined for the domain bounded by

$$x=0, \quad x=1, \quad y=0, \quad y=1,$$

by the rule that, for every rational value of x of the form

$$\frac{2m+1}{2^n} \cdot (n \geqq 0), \quad \psi(x) = \frac{1}{2^n},$$

and that, for every other value of x, $\psi(x) = 0$.

Let $f(x, y) = \left| \frac{1}{y} \sin \frac{1}{y} \right| \psi(x)$, then the improper integral

$$\int \left| \frac{1}{y} \sin \frac{1}{y} \right| \psi(x)\, (dx\, dy)$$

exists as a Jordan double integral, and has the value zero. The integral $\int_0^1 \psi(x)\, dx$ exists as a Riemann integral, and has the value zero. The repeated integral

$$\int_0^1 dx \int_0^1 \psi(x) \left| \frac{1}{y} \sin \frac{1}{y} \right| dy$$

does not exist in accordance with Harnack's definition, for

$$\int_0^1 \psi(x) \left| \frac{1}{y} \sin \frac{1}{y} \right| dy$$

diverges for the everywhere-dense set of values $x = \frac{2m+1}{2^n}$, and therefore the repeated

* Stolz, *Grundzüge*, vol. III, p. 149. The function $\psi(x)$ was first given by Du Bois Reymond, *Crelle's Journal*, vol. xcvi, p. 278.

integral does not exist. It does exist, however, in accordance with Lebesgue's definition, since the points of divergence of the single integral form a measurable set of measure zero; and its value is zero, the same as that of the double integral. The other repeated integral exists, and is zero.

5. Let $f(x, y)=0$*, at all points in the rectangle bounded by $x=0$, $x=1$, $y=0$, $y=1$; except that at all points $x=\frac{2m+1}{2^s}$, $y\leq\frac{1}{2^s}$, it is to have the improper value $+\infty$. The function $f_n(x, y)$ will be given by the condition that it equals N_n, at the points where $f(x, y)=\infty$, and is elsewhere zero. It can be shewn that

$$\int f_n(x, y)\,(dx\,dy)=0,$$

for every value of N_n; and thus $\int f(x, y)\,(dx dy)$ exists, and $=0$. The condition

$$\int_0^1 f(x, y)\,dy-\int_0^1 f_n(x, y)\,dy<\epsilon$$

is not satisfied for any of the everywhere-dense set of values $x=(2m+1)/2^n$, if the ordinary definitions of Harnack and Riemann are employed; and therefore the convergence of $\int f_n(x, y)\,dy$ to $\int f(x, y)\,dy$ is not regular. The repeated integral

$$\int_0^1 dx\int_0^1 f(x, y)\,dy$$

is equivalent to $\int_0^1 dx \lim\int_0^1 f_n(x, y)\,dy$. Also $\int_0^1 f_n(x, y)\,dy$ is zero, unless

$$x=(2m+1)/2^s,$$

in which case it is $N_n/2^s$, and for such values of x, $\lim\int_0^1 f_n(x, y)\,dy$ is ∞; and thus the repeated integral does not exist, in accordance with Harnack's definition. It does, however, exist in accordance with Lebesgue's definition, and is equal to zero, the value of the double integral.

REPEATED INTEGRALS OVER AN INFINITE DOMAIN.

409. Let the function $f(x, y)$ be defined for all points of the infinite domain of which the boundaries are the three straight lines $x=a$, $x=b$, $y=c$; the domain being unbounded in the positive direction of y. Conditions will be investigated that the repeated integrals

$$\int_a^b dx\int_c^\infty f(x, y)\,dy,\quad \int_c^\infty dy\int_a^b f(x, y)\,dx,$$

may exist, and may be equal to one another.

* This example is given by Schönflies, *Bericht*, pp. 201, 202, to illustrate his erroneous theorem, applicable to Jordan and Harnack integrals, that the improper double integral is always equal to each of the repeated integrals, and the converse; and that thus the condition of the theorem of § 400 is always satisfied. The example does not bear out his contention. See Hobson, *Proc. Lond. Math. Soc.*, ser. 2, vol. IV, pp. 157—159.

Let it be assumed that the two repeated integrals

$$\int_a^b dx \int_c^Y f(x, y)\, dy, \quad \int_c^Y dy \int_a^b f(x, y)\, dx$$

exist, and are equal, for every definite value of $Y \geqq c$.

If $\int_c^\infty dy \int_a^b f(x, y)\, dx$ exist as a definite number, then it is equal to

$$\lim_{Y=\infty} \int_a^b dx \int_c^Y f(x, y)\, dy;$$

and this is equal to

$$\int_a^b dx \int_c^\infty f(x, y)\, dy,$$

provided that

$$\lim_{Y=\infty} \int_a^b dx \int_Y^\infty f(x, y)\, dy = 0.$$

This last condition is equivalent to the condition that, corresponding to an arbitrarily chosen positive ϵ, a value Y_ϵ of Y can be so chosen that

$$\left| \int_a^b dx \int_Y^\infty f(x, y)\, dy \right| < \epsilon, \text{ for } Y > Y_\epsilon.$$

We can now state the following result:—

On the supposition that the two repeated integrals of $f(x, y)$ both exist, and have equal values in the domain bounded by $x = a$, $x = b$, $y = c$, $y = Y$, for every value of $Y \geqq c$, the necessary and sufficient conditions that the repeated integrals of $f(x, y)$ in the unbounded domain $x = a$, $x = b$, $y = c$, $y = \infty$, shall have equal values are

(1) *that $\int_c^\infty dy \int_a^b f(x, y)\, dx$ shall have a definite value; and*

(2) *that, corresponding to an arbitrarily chosen ϵ, a value Y_ϵ of Y can be fixed, so that $\left| \int_a^b dx \int_Y^\infty f(x, y)\, dy \right| < \epsilon$, for all values of $Y > Y_\epsilon$.*

It may be observed that the condition (2) does not make it necessary that $\int_Y^\infty f(x, y)\, dy$ should have a definite value for all values of x.

It is a sufficient condition in order that (2) may be satisfied, that

$$\left| \int_Y^\infty f(x, y)\, dy \right| < \epsilon/(b - a),$$

for every value of x in (a, b), with the possible exception of a set of points of zero measure.

In this exceptional case, however, $\int_a^b dx \int_Y^\infty f(x, y)\, dy$, must exist as an absolutely convergent integral with respect to x.

410. Next, let us assume the function $f(x, y)$, to be defined for the infinite domain bounded by $x = a$, $y = c$, and unbounded in the positive directions of x and y.

Let it be assumed that the two repeated integrals

$$\int_a^X dx \int_c^Y f(x, y)\, dy, \quad \int_c^Y dy \int_a^X f(x, y)\, dx$$

exist, and have the same value, for each set of definite values of X and Y, such that $X \geqq a$, $Y \geqq b$. The value of these repeated integrals we denote by $\phi(X, Y)$.

We have now

$$\lim_{Y=\infty} \phi(X, Y) = \int_c^\infty dy \int_a^X f(x, y)\, dx = \lim_{Y=\infty} \int_a^X dx \int_c^Y f(x, y)\, dy;$$

it being assumed that this limit has a definite value for each definite value of X. If now

$$\lim_{Y=\infty} \int_a^X dx \int_c^Y f(x, y)\, dy = \int_a^X dx \lim_{Y=\infty} \int_c^Y f(x, y)\, dy,$$

we then have

$$\lim_{Y=\infty} \phi(X, Y) = \int_a^X dx \int_c^\infty f(x, y)\, dy.$$

The condition that this may be the case is that

$$\lim_{Y=\infty} \int_a^X dx \int_Y^\infty f(x, y)\, dy = 0,$$

which is equivalent to the condition that, corresponding to an arbitrarily chosen ϵ, there should exist, for each value of X, a value Y_ϵ of Y, such that

$$\left| \int_a^X dx \int_{Y_\epsilon}^Y f(x, y)\, dy \right| < \epsilon,$$

for every value of $Y > Y_\epsilon$. When this condition is satisfied, we have

$$\lim_{X=\infty} \lim_{Y=\infty} \phi(X, Y) = \int_0^\infty dx \int_0^\infty f(x, y)\, dy,$$

provided the repeated limit on the left-hand side has a definite value.

Under similar conditions we can shew, in the same manner, that

$$\lim_{Y=\infty} \lim_{X=\infty} \phi(X, Y) = \int_0^\infty dy \int_0^\infty f(x, y)\, dx.$$

The necessary and sufficient condition for the existence and equality of the repeated limits $\lim\limits_{X=\infty} \lim\limits_{Y=\infty} \phi(X, Y)$, $\lim\limits_{Y=\infty} \lim\limits_{X=\infty} \phi(X, Y)$ is obtained from the theorem of § 234, by letting $X = 1/x$, $Y = 1/y$, where x, y are the pair of variables there employed. Using this condition, we have now established the following theorem :—

It being assumed that the repeated integrals of $f(x, y)$ in a domain bounded by $x=a$, $x=X$, $y=c$, $y=Y$, exist, and are equal to one another, for every pair of definite values of X and Y such that $X \geqq a$, $Y \geqq c$, it is sufficient for the existence and equality of the two repeated integrals*

$$\int_a^\infty dx \int_c^\infty f(x, y)\, dy, \quad \int_c^\infty dy \int_a^\infty f(x, y)\, dx,$$

that the following conditions be satisfied:—

(1) *That* $\int_c^\infty dy \int_a^X f(x, y)\, dx$, $\int_a^\infty dx \int_c^Y f(x, y)\, dy$ *have definite values for all definite values of X and Y.*

(2) *That, corresponding to an arbitrary ϵ, there should exist, for each value of X, a value Y_ϵ of Y, such that*

$$\left| \int_a^X dx \int_{Y_\epsilon}^Y f(x, y)\, dy \right| < \epsilon, \text{ for every } Y > Y_\epsilon;$$

and also for each value of Y, a value X_ϵ of X, such that

$$\left| \int_c^Y dy \int_{X_\epsilon}^X f(x, y)\, dx \right| < \epsilon, \text{ for every } X > X_\epsilon.$$

(3) *That if ϵ be fixed, a number $Y_0 > c$ can be determined, such that, for each value of $Y > Y_0$, a value of X, say X_ϵ, can be found, which is such that, for the particular value of Y,*

$$\left| \int_a^X dx \int_Y^\infty f(x, y)\, dy \right| < \epsilon,$$

for every value of $X > X_\epsilon$. This condition may be replaced by the corresponding one in which the integral

$$\int_c^Y dy \int_X^\infty f(x, y)\, dx \text{ is employed.}$$

More general conditions could be obtained by not assuming that

$$\lim_{Y=\infty} \phi(X, Y), \qquad \lim_{X=\infty} \phi(X, Y)$$

have necessarily definite meanings, and then applying, instead of (3), the theorem of § 233.

In accordance with § 233, if, in condition (3), when X_ϵ is determined for a particular value of Y, the condition be satisfied, not only for that particular value of Y, but also for all greater values, then the double limit

$$\lim_{X=\infty,\, Y=\infty} \int_a^X dx \int_c^Y f(x, y)\, dy$$

also exists, and is equal to the repeated integrals with infinite limits.

* Conditions nearly equivalent to these were given by Bromwich, *Proc. Lond. Math. Soc.*, ser. 2, vol. I, p. 188.

It has been proposed by Hardy* to employ this double limit as a definition of the double integral of $f(x, y)$ over the infinite domain. This definition is, of course, less stringent than the one given in § 323.

411. Sufficient conditions of greater stringency, and therefore of less wide application, can be deduced from the theorem of § 410.

The integral $\int_a^\infty f(x, y)\,dx$, is said to *converge uniformly* in *the interval* (c, Y), if, having given the positive number η, chosen arbitrarily, it be always possible to determine X_0 so that

$$\left|\int_X^\infty f(x, y)\,dx\right| < \eta,$$

for $X \geqq X_0$, and for every value of y in the interval (c, Y).

If this condition be satisfied for each interval (c, Y) of y, the convergence is said to be *uniform in an arbitrary interval.*

If the condition be satisfied for every value of $Y > c$, the convergence is said to be uniform in an unlimited interval.

If the condition be satisfied for all points in (c, Y) except those which belong to a set of points of zero measure, the convergence is said to be *uniform in general* in the interval.

The following theorem† will be established:—

It is sufficient for the existence and equality of the repeated integrals with infinite limits, (1) *that the repeated integrals between finite limits always exist and are equal,* (2) *that* $\int_a^\infty f(x, y)\,dx$ *be uniformly convergent in general in an arbitrary interval,* (3) *that* $\int_c^\infty f(x, y)\,dy$ *satisfy the same condition, and* (4) *that* $\int_a^\infty dx \int_c^Y f(x, y)\,dy$ *converge uniformly in an unlimited interval. The condition of uniform convergence in general must be replaced by that of uniform convergence, in case the repeated integrals between finite limits exist only as non-absolutely convergent improper integrals.*

To prove this theorem it will be sufficient to shew that if the conditions given in it be all satisfied, then those in the theorem of § 410 are all satisfied.

* *Messenger of Math.*, vol. XXXII, p. 95.

† This theorem was given by de la Vallée-Poussin, *Liouville's Journal*, ser. 4, vol. VIII, p. 464, except that in his definition of uniform convergence in general, the set of exceptional points is there restricted to be a set of the first species; also the condition (2) is given as that of uniform convergence.

It is clear that the condition (2) of § 410 is satisfied, if (2) and (3) of the present theorem are satisfied. For, if $\int_c^\infty f(x, y)\,dy$ be uniformly convergent in general in the interval (x, X), then for a fixed η, Y_0 can be so fixed that

$$\left|\int_Y^\infty f(x, y)\,dy\right| < \eta,$$

for $Y \geqq Y_0$, and for every value of x in (a, X) except the points of a set of zero measure. It then follows that

$$\left|\int_a^X dx \int_Y^\infty f(x, y)\,dy\right| < \eta\,(X - a),$$

and we may choose $\eta < \epsilon/2(X - a)$; therefore

$$\left|\int_a^X dx \int_{Y_0}^Y f(x, y)\,dy\right| < \epsilon,$$

for $Y \geqq Y_0$; this is one part of condition (2) of § 410; similarly it may be shewn that (2) of the present theorem is sufficient to satisfy the other part of the condition (2) of the former theorem. Again, the condition (4) of the present theorem may be stated in the form that

$$|\,\phi(X, Y) - \lim_{X=\infty} \phi(X, Y)\,| < \epsilon,$$

for every value of Y, and for all values of X which are $\geqq$ a fixed value X_0. Since, on account of (2),

$$\lim_{X=\infty} \phi(X, Y) = \int_c^Y dy \int_a^\infty f(x, y)\,dx,$$

we see that

$$\left|\int_c^Y dy \int_X^\infty f(x, y)\,dx\right| < \epsilon,$$

for every Y, and for $X \geqq X_0$, and therefore condition (3) of the former theorem is satisfied. It has thus been shewn that, if the conditions of the theorem are satisfied, then those of the theorem in § 410 are also satisfied.

EXAMPLES.

1. It will be found that $\int_1^\infty dx \int_1^\infty \frac{x^2 - y^2}{(x^2 + y^2)^2}\,dy = -\frac{1}{4}\pi$, and that

$$\int_1^\infty dy \int_1^\infty \frac{x^2 - y^2}{(x^2 + y^2)^2}\,dx = \frac{1}{4}\pi.$$

In this case the first condition of the theorem in § 410 is satisfied, and the second is also satisfied. For

$$\int_1^X dx \int_{Y_\epsilon}^Y \frac{x^2 - y^2}{(x^2 + y^2)^2}\,dy = \left\{\tan^{-1}\frac{X}{Y} - \tan^{-1}\frac{1}{Y}\right\} - \left\{\tan^{-1}\frac{X}{Y_\epsilon} - \tan^{-1}\frac{1}{Y_\epsilon}\right\};$$

and hence it is clear that, for a given ϵ, Y_ϵ can be so determined that, for every value of $Y > Y_\epsilon$, the absolute value of the repeated integral is less than ϵ. The third condition is, however, not satisfied; for, since

$$\int_1^X dx \int_Y^\infty \frac{x^2-y^2}{(x^2+y^2)^2}\, dy = \tan^{-1}\frac{1}{Y} - \tan^{-1}\frac{X}{Y},$$

it is impossible, when Y is fixed, so to choose X_ϵ, that, for every $X > X_\epsilon$, the absolute value of the repeated integral shall be less than ϵ.

2. Let* $\phi(z) = \dfrac{pz^p}{1+z^{2p}}$, where $p>1$; then $\phi'(z) = \dfrac{p^2 z^{p-1}(1-z^{2p})}{(1+z^{2p})^2}$.

The repeated integral $\int_0^c dy \int_0^\infty \phi'(xy)\, dx = 0$, but $\int_0^\infty dx \int_0^c \phi'(xy)\, dy = \frac{1}{2}\pi$. The second condition of the theorem in § 410 is here not satisfied. For, we find

$$\int_0^c dy \int_X^\infty \phi'(xy)\, dx = -\tan^{-1}(c^p X^p),$$

and it is impossible to choose X_ϵ so that, for all values of $X > X_\epsilon$, the absolute value of this repeated integral shall be less than ϵ.

3. We have $\int_0^\infty dy \int_0^a \cos xy\, dx = \frac{1}{2}\pi$, for $a>0$; but $\int_0^a dx \int_0^\infty \cos xy\, dy$ does not exist.

4. It may be shewn that $\int_0^\infty dx \int_0^c e^{-xy} dy = \int_0^c dy \int_0^\infty e^{-xy} dx$, the conditions of the theorem in § 409 being satisfied.

5. Let† $V = \left(\dfrac{1}{x} + \dfrac{1}{y}\right) \sin \pi x \sin \pi y$. In this case we find

$$\int_1^h dx \int_1^k \frac{\partial^2 V}{\partial x \partial y}\, dy = \int_1^k dy \int_1^h \frac{\partial^2 V}{\partial x \partial y}\, dx = \left(\frac{1}{h} + \frac{1}{k}\right) \sin \pi h \sin \pi k.$$

The repeated integral $\int_1^\infty dx \int_1^\infty \frac{\partial^2 V}{\partial x \partial y}\, dy$ does not however exist; for $\int_1^\infty \frac{\partial^2 V}{\partial x \partial y}\, dy$ or $\left[\frac{\partial V}{\partial x}\right]_1^\infty$ has no definite value, for any value of x. The double limit

$$\lim_{h=\infty, k=\infty} \int_1^h dx \int_1^k \frac{\partial^2 V}{\partial x \partial y}\, dy$$

exists, and is equal to zero.

In this case, as stated in § 410, it is possible to employ this double limit as the definition of the double integral.

6. Let $V = \left(\dfrac{1}{x} + \dfrac{1}{y}\right)\left(1 - \dfrac{1}{y}\right) \sin \pi x$, and $f(x, y) = \dfrac{\partial^2 V}{\partial x \partial y}$. We find, in this case,

$$\int_1^\infty dx \int_1^\infty f(x, y)\, dy = 0,$$

but the other repeated integral does not exist, since $\int_1^\infty f(x, y)\, dx$ has no definite value. The double limit

$$\lim_{h=\infty,\, k=\infty} \int_1^h dx \int_1^k f(x, y)\, dy = 0.$$

* See Stolz, *Grundzüge*, vol. III, pp. 8, 182, where the example is ascribed to Du Bois Reymond.

† Bromwich, *Proc. Lond. Math. Soc.*, ser. 2, vol. I, p. 182.

7. Let $f(x, y)=\dfrac{\partial^2 V}{\partial x \partial y}$, $V=\dfrac{xy}{1+x^2+y^2}$. In this case, the two repeated integrals

$$\int_0^\infty dx \int_0^\infty f(x, y)\, dy, \quad \int_0^\infty dy \int_0^\infty f(x, y)\, dx$$

exist and are zero.

The conditions of the theorem of § 410 are here satisfied. We find that

$$\int_0^\infty dy \int_0^X f(x, y)\, dx = 0 = \int_0^\infty dx \int_0^Y f(x, y)\, dy\, ;$$

and thus the first condition is satisfied. Again

$$\int_0^X dy \int_{Y_\epsilon}^Y f(x, y)\, dy = \frac{XY}{1+X^2+Y^2} - \frac{XY_\epsilon}{1+X^2+Y_\epsilon^2}\, ;$$

and, for a given X, this is arbitrarily small for $Y > Y_\epsilon$, provided Y_ϵ be properly chosen. Since $\int_0^X dx \int_Y^\infty f(x, y)\, dy = \dfrac{-XY}{1+X^2+Y^2}$, it can be at once verified that the third condition is satisfied. It will be found that the conditions of the theorem in § 411 are not satisfied in this case.

8. To prove that the order of integration of

$$\int_0^\infty \sin y\, dy \int_0^\infty e^{-yx^2}\, dx$$

may be reversed, the theorem of § 410 must be applied.

Let $\phi(\xi) = \int_\xi^\infty e^{-\xi^2} d\xi$; then $\phi(0) = \frac{1}{2}\sqrt{\pi}$, and $\phi(\xi)$ continually diminishes as ξ increases. We have

$$\int_0^\infty dy \int_0^X e^{-yx^2} \sin y\, dx = \int_0^\infty \frac{\sin y}{\sqrt{y}} [\phi(0) - \phi(X\sqrt{y})]\, dy,$$

and this is easily seen to be less than

$$\phi(0) \int_0^\pi \frac{\sin y}{\sqrt{y}}\, dy\, ;$$

therefore the repeated integral exists. Again

$$\int_0^\infty dx \int_0^Y e^{-yx^2} \sin y\, dy = \int_0^\infty \frac{1}{x^4+1} [1 - e^{-x^2 Y}(\cos Y + x^2 \sin Y)]\, dx,$$

and this is easily seen to be convergent; therefore condition (1) is satisfied.

Next, we have

$$\int_{Y_\epsilon}^Y e^{-yx^2} \sin y\, dy = \frac{1}{1+x^4} [e^{-Y_\epsilon x^2}(\cos Y_\epsilon + x^2 \sin Y_\epsilon) - e^{-Yx^2}(\cos Y + x^2 \sin Y)],$$

and this is less than $2(1+x^2)e^{-Y_\epsilon x^2}$, where $Y > Y_\epsilon$. We then find that

$$\int_0^X dx \int_{Y_\epsilon}^Y e^{-yx^2} \sin y\, dy$$

is numerically less than

$$\sqrt{\frac{\pi}{Y_\epsilon}} \left(1 + \frac{1}{2Y_\epsilon}\right);$$

and it is then seen that condition (3), and the first part of (2) are satisfied.

Also

$$\int_0^Y dy \int_{X_\epsilon}^X e^{-yx^2} \sin y\, dx = \int_0^Y \frac{\sin y}{\sqrt{y}} [\phi(X_\epsilon \sqrt{y}) - \phi(X\sqrt{y})]\, dy;$$

since $\phi(\xi) < \frac{1}{2\xi}$, we have

$$\int_0^Y \frac{\sin y}{\sqrt{y}} \phi(X_\epsilon \sqrt{y})\, dy < \int_0^\pi \frac{\sin y}{\sqrt{y}} \phi(X_\epsilon \sqrt{y})\, dy < \frac{\pi}{4X_\epsilon};$$

and it is now clear that the second part of (2) is satisfied.

THE LIMIT OF AN INTEGRAL CONTAINING A PARAMETER.

412. When an integral $\int_a^b f(x, y)\, dx$, involving a parameter y, exists for each value of y such that $y_0 < y \leqq y_0 + h$, it is of importance to know under what circumstances the limit of the value of the integral as y converges to y_0, has a definite value which is obtained by integrating through (a, b) the limit of $f(x, y)$ for $y = y_0$. Expressed symbolically, and denoting $\lim_{y=y_0} f(x, y)$ by $f(x, y_0 + 0)$, we require conditions under which $\lim_{y=y_0} \int_a^b f(x, y)\, dx$, and $\int_a^b f(x, y_0 + 0)\, dx$ may both exist, and have one and the same value.

In order that $\int_a^b f(x, y)\, dx$ may be continuous at $y = y_0$, on the right, it is necessary that the further condition be satisfied that $f(x, y_0 + 0) - f(x, y)$ should be an integrable null-function in (a, b). In particular, if

$$f(x, y) = f(x, y_0 + 0),$$

for all values of x in (a, b), this condition is satisfied.

If $y_1, y_2, \ldots y_n, \ldots$ be a sequence of diminishing values of y which converges to the limit y_0, we may write $s_n(x)$ for $f(x, y_n)$, and $s(x)$ for $f(x, y_0 + 0)$, and then the results of §§ 383–386, may be applied to obtain sufficient conditions that, for this sequence of values of y, the expressions

$$\int_a^b f(x, y_0 + 0)\, dx, \quad \lim_{y=y_0} \int_a^b f(x, y)\, dx,$$

may exist, and be equal, when the particular sequence of values of y is alone considered.

Criteria of the required character will be obtained by adding the condition that the sufficient conditions so obtained are satisfied for every possible sequence of diminishing values of y which converges to y_0. It is clear that the case in which h is negative, and $f(x, y_0 - 0)$ takes the place of $f(x, y_0 + 0)$, is not essentially different from the case here considered.

It is clear from the result of § 383, that, in case $f(x, y)$ converges to $f(x, y_0 + 0)$ uniformly for all values of x in the interval (a, b), sufficient conditions for the equality in question are satisfied for every sequence of values of y which converges to y_0. We have therefore the following theorem:—

When the proper integral $\int_a^b f(x, y)\,dx$ *exists, for each value of* y *such that* $y_0 < y \leqq y_0 + h$, *it is a sufficient condition that* $\int_a^b f(x, y_0 + 0)\,dx$ *may exist and be equal to* $\lim_{y=y_0} \int_a^b f(x, y)\,dx$, *that* $f(x, y)$ *should converge to* $f(x, y_0 + 0)$ *uniformly for all values of* x *in the interval* (a, b).

The condition of uniform convergence is that, corresponding to an arbitrarily chosen positive number ϵ, a number k can be determined, such that $|f(x, y + \delta) - f(x, y_0 + 0)| < \epsilon$, provided $0 < \delta < k$, for every value of x in (a, b).

We may obtain less stringent conditions than the one contained in the above theorem by applying other results relating to the integration of series.

From the theorem of § 380, we obtain the following criterion for the existence of $\int_a^b f(x, y_0 + 0)\,dx$:—

If $\int_a^b f(x, y)\,dx$ *be a proper integral, for each value of* y *such that* $y_0 < y \leqq y_0 + h$, *and* $f(x, y_0 + 0)$ *be limited in the interval* (a, b), *the necessary and sufficient condition that* $\int_a^b f(x, y_0 + 0)\,dx$ *should exist as a Riemann integral is that* $f(x, y)$ *may converge to* $f(x, y_0 + 0)$ *regularly in* (a, b), *except for a set of points* E, *of zero measure, and of the first category.*

The condition of regular convergence, except for the set E, is that, if ϵ be an arbitrarily chosen positive number, and y_1 an arbitrarily chosen value of y such that $y_0 < y_1 \leqq y_0 + h$, then for every point x_1 which does not belong to a certain non-dense closed component of E, dependent on ϵ, a value y_{x_1} of $y\,(< y_1)$ can be chosen, and also an interval $(x_1 - \delta, x_1 + \delta')$ can be fixed, such that $|f(x, y_{x_1}) - f(x, y_0 + 0)| < \epsilon$, for every value of x belonging to (a, b) in $(x_1 - \delta, x_1 + \delta')$.

When the conditions stated in the above theorem for the existence of $\int_a^b f(x, y_0 + 0)\,dx$ are satisfied, we may apply the results of § 383 and § 385, to obtain the following theorems:—

If $\int_a^b f(x, y)\,dx$ *be a proper integral, for each value of* y *such that* $y_0 < y \leqq y_0 + h$, *and the proper integral* $\int_a^b f(x, y_0 + 0)\,dx$ *also exist, it is a* *sufficient condition that* $\lim_{y=y_0}\int_a^b f(x, y)\, dx$ *should exist and be equal to* $\int_a^b f(x, y_0+0)\, dx$, *that* $|f(x, y)|$ *should be less than some fixed finite number, for all values of* x *and* y *such that* $a \leqq x \leqq b$, *and* $y_0 < y \leqq y_0 + h$.

If $\int_a^b f(x, y)\, dx$ *be a proper integral for each value of* y *such that* $y_0 < y \leqq y_0 + h$, *and the proper integral* $\int_a^b f(x, y_0+0)\, dx$ *exist, it is sufficient for the equality of*

$$\lim_{y=y_0}\int_a^b f(x, y)\, dx \text{ and } \int_a^b f(x, y_0+0)\, dx,$$

that (1) $\lim_{y=y_0}\int_a^x f(x, y)\, dx$ *should be a continuous function of* x *in the whole interval* (a, b), *and that* (2) *the set of points* (x, y_0) *at which the saltus of* $f(x, y)$ *considered as a function of* (x, y), *is indefinitely great, should form at most an enumerable set.*

413. In the case in which the integrals $\int_a^b f(x, y)\, dx$, for values of y such that $y_0 < y \leqq y_0 + h$, are not necessarily proper integrals, but may, for some or all such values of y, exist only as improper integrals, we may apply the result of § 386, to obtain a set of sufficient conditions for the equality of $\int_a^b f(x, y_0+0)\, dx$, and $\lim_{y=y_0}\int_a^b f(x, y)\, dx$.

The result is stated in the following theorem:—

If $f(x, y)$ *converges to a definite limit* $f(x, y_0+0)$ *for all points* x *of the interval* (a, b) *which do not belong to a reducible set of points* G, *and the functions* $f(x, y)$, *for* $y_0 < y < y_0 + h$, *although not necessarily limited in* (a, b), *satisfy the conditions* (1) *that, in any interval* (α, β) *contained in* (a, b) *which contains in its interior and at its ends no point of* G, $|f(x, y)|$ *is less than some fixed finite number, for all values of* x *and* y *such that* $\alpha \leqq x \leqq \beta$, $y_0 < y \leqq y_0 + h$; (2) *that* $\int_a^b f(x, y)\, dx$ *exists at least as an improper integral, for each value of* y *such that* $y_0 < y < y_0 + h$; *and* (3) *that* $\lim_{y=y_0}\int_a^x f(x, y)\, dx$, *for* $a \leqq x \leqq b$ *is convergent and represents a continuous function of* x; *and* (4) $\int_a^b f(x, y_0+0)\, dx$ *exists at least as an improper integral; then the equality* $\int_a^b f(x, y_0+0)\, dx = \lim_{y=y_0}\int_a^b f(x, y)\, dx$ *holds.*

414. In the case in which the interval of integration is unlimited, say $b = \infty$, we may apply the theorems of § 387, to obtain sufficient conditions that $\int_a^\infty f(x, y_0+0)\,dx = \lim_{y=y_0}\int_a^b f(x, y)\,dx$. We obtain at once the following criteria:—

If the equality $\int_a^C f(x, y_0+0)\,dx = \lim_{y=y_0}\int_a^C f(x, y)\,dx$ *be satisfied for every value of* $C > a$, *then if, corresponding to an arbitrarily fixed* ϵ, *a number* $C > a$, *can be determined, and also a value* y_1 *of* y, *such that* $y_0 < y_1 < y_0 + h$, *for which* $\left|\int_C^{C'} f(x, y)\,dx\right| < \epsilon$, *for every value of* $C' > C$, *and for every value of* y *such that* $y_0 < y \leqq y_1$, *then* $\int_a^\infty f(x, y_0+0)\,dx$ *exists, and is equal to* $\lim_{y=y_0}\int_a^\infty f(x, y)\,dx$.

If the equality $\int_a^C f(x, y_0+0)\,dx = \lim_{y=y_0}\int_a^C f(x, y)\,dx$ *holds for every value of* $C > a$, *then, if* $\lim_{y=y_0}\int_a^\infty f(x, y)\,dx$ *exists, and also* $\lim_{y=y_0}\int_a^C f(x, y)\,dx$ *converges to the value* $\lim_{y=y_0}\int_a^\infty f(x, y)\,dx$, *when* C *is indefinitely increased, these conditions are sufficient to ensure that* $\int_a^\infty f(x, y_0+0)\,dx$ *exists, and is equal to* $\lim_{y=y_0}\int_a^\infty f(x, y)\,dx$.

In order that $\int_a^\infty f(x, y)\,dx$ may be continuous on the right at y_0, the additional condition must be satisfied that $f(x, y_0+0) = f(x, y_0)$; or more generally, that $f(x, y_0+0) - f(x, y_0)$ should be an integrable null-function in an arbitrary interval of x.

EXAMPLES.

1. If $y>0$, we have $\int_0^\infty \frac{\sin yx}{x}\,dx = \frac{1}{2}\pi$, but when $y=0$, $\int_0^\infty \frac{\sin yx}{x}\,dx$ vanishes; and thus $\int_0^\infty \frac{\sin yx}{x}\,dx$ is discontinuous, for $y=+0$. It may be seen that the conditions contained in neither of the above theorems are satisfied.

The condition $\lim_{y=0}\int_0^C \frac{\sin yx}{x}\,dx = \int_0^C \lim_{y=0}\frac{\sin yx}{x}\,dx = 0$ is satisfied. But

$$\int_C^{C'} \frac{\sin yx}{x}\,dx = \int_{Cy}^{C'y} \frac{\sin\theta}{\theta}\,d\theta > 1 - \frac{2}{\pi}Cy, \text{ if } C'y = \tfrac{1}{2}\pi.$$

However y_1 and C may be fixed, C' and $y\,(<y_1)$ can always be so chosen that $\left|\int_C^{C'} \frac{\sin yx}{x}\,dx\right| > \epsilon$; and thus the second condition is not satisfied.

2. The equality

$$\lim_{y=y_0}\int_a^b f(x, y)\,\phi(x)\,dx = \int_a^b \phi(x) f(x, y_0+0)\,dx$$

holds if $|f(x, y)|$ is less than some fixed number, for all values of x and y such that

$a \leqq x \leqq b$, $y_0 < y \leqq y_0 + h$; provided also $\int_a^b \phi(x)\,dx$ is either a proper integral, or an absolutely convergent improper integral with a reducible set of points of infinite discontinuity. This follows from the theorem in § 413. We may even suppose $b = \infty$; then under the same conditions the equality holds. For

$$\left| \int_C^{C'} f(x, y)\, \phi(x)\, dx \right|$$

is less than $K \int_C^{C'} \phi(x)\, dx$, where K is the upper limit of $|f(x, y)|$, and therefore for a sufficiently great value of C, we have

$$\left| \int_C^{C'} f(x, y)\, \phi(x)\, dx \right| < \epsilon,$$

since C may be so chosen that

$$\left| \int_C^{C'} \phi(x)\, dx \right| < \frac{\epsilon}{K}.$$

It follows that

$$\int_a^\infty f(x, y_0 + 0)\, \phi(x)\, dx$$

exists, and is equal to

$$\lim_{y = y_0} \int_a^\infty f(x, y)\, \phi(x)\, dx,$$

provided $|f(x, y)|$ is less than K, and also provided $\int_a^\infty \phi(x)\, dx$ is absolutely convergent.

3. Consider $\int_a^b e^{-yx} \phi(x)\, dx$, where a is finite, and b is either finite, or $+\infty$. It follows from Ex. (2), that

$$\lim_{y=0} \int_a^b e^{-yx} \phi(x)\, dx = \int_a^b \phi(x)\, dx,$$

provided $\int_a^b \phi(x)\, dx$ is absolutely convergent, and $\phi(x)$ has at most a reducible set of points of infinite discontinuity. The theorem however holds whenever $\int_a^b e^{-yx} \phi(x)\, dx$ has a definite value not only for $y = 0$, but also for all values of y such that $0 \leqq y \leqq h$, where h is some positive number. If $\psi(x)$ denote the continuous function $\int_a^x \phi(x)\, dx$, we have

$$\int_a^b e^{-yx} \phi(x)\, dx = e^{-by} \psi(b) + y \int_a^b e^{-yx} \psi(x)\, dx,$$

b being taken to be finite. Since $|\psi(x)|$ has a finite upper limit U in (a, b), we have

$$\left| \int_a^b e^{-yx} \psi(x)\, dx \right| < U e^{-ay} (b - a);$$

therefore

$$\lim_{y=0} \int_a^b e^{-yx} \phi(x)\, dx = \psi(b) = \int_a^b \phi(x)\, dx.$$

In case $b = \infty$, we have

$$\int_a^\infty e^{-yx} \phi(x)\, dx = y \int_a^\infty e^{-yx} \psi(x)\, dx$$

$$= y \int_a^{\frac{1}{\sqrt{y}}} e^{-yx} \psi(x)\, dx + y \int_{\frac{1}{\sqrt{y}}}^\infty e^{-yx} \psi(x)\, dx.$$

Hence, by applying the first mean value theorem, we have

$$\int_0^\infty e^{-yx} \phi(x)\, dx = \psi(x_1)\,(e^{-ay} - e^{-\sqrt{y}}) + \psi(x_2)\, e^{-\sqrt{y}}$$

where x_1 is some number between a and $1/\sqrt{y}$, and x_2 some number greater than $1/\sqrt{y}$. When y converges to the limit zero, the first term on the right-hand side converges to the limit zero, and the second term to the limit $\psi(\infty)$, or $\int_a^\infty \phi(x)\,dx$. Thus the theorem is established for the case $b=\infty$.

DIFFERENTIATION OF AN INTEGRAL WITH RESPECT TO A PARAMETER.

415. Let $f(x, y)$ be a function of the variable x and the parameter y, and which we shall suppose to be defined in the domain of (x, y) defined by $a \leqq x \leqq b$, $y_0 \leqq y \leqq y_0 + \alpha$. Further, let the integral $\int_a^b f(x, y)\,dx$ exist as a proper integral, for each value of y in the interval $(y_0, y_0+\alpha)$. If u denote $\int_a^b f(x, y)\,dx$, sufficient conditions will be investigated that

$$\left(\frac{\partial u}{\partial y}\right)_{y=y_0} = \int_a^b \left\{\frac{\partial f(x, y)}{\partial y}\right\}_{y=y_0} dx.$$

This is the ordinary rule first employed by Leibnitz, of differentiation under the sign of integration, and is an important example of the process of changing the order of repeated limits.

In this result $\left(\frac{\partial u}{\partial y}\right)_{y=y_0}$ denotes the derivative at u on one side. No information is afforded as to the existence of a derivative on the other side. If, however, the function be defined for values of y on the other side of y_0, the derivative on that side may be treated in a similar manner.

We have

$$\frac{u_{y_0+h} - u_{y_0}}{h} = \int_a^b \frac{f(x, y_0+h) - f(x, y_0)}{h}\,dx,$$

where $h < \alpha$. If it now be assumed that, for each value of x, $f(x, y)$ has a differential coefficient with respect to y, at all points interior to the interval $(y_0, y_0+\alpha)$ of y, we have from the theorem of § 203,

$$\frac{u_{y_0+h} - u_{y_0}}{h} = \int_a^b \frac{\partial f(x, y_0+\theta h)}{\partial y}\,dx,$$

where $0 < \theta < 1$, and the value of θ depends in general upon x and h.

If it be now further assumed that $\frac{\partial f(x, y)}{\partial y}$ converges to $\left\{\frac{\partial f(x, y)}{\partial y}\right\}_{y=y_0}$ as y converges to y_0, uniformly for the interval (a, b) of x, we have, for a sufficiently small value of h,

$$\frac{\partial f(x, y_0+\theta h)}{\partial y} = \left\{\frac{\partial f(x, y)}{\partial y}\right\}_{y=y_0} + \beta(x),$$

where $|\beta(x)| < \epsilon$, for every value of x.

Under these conditions, we have, since $\left|\int_a^b \beta(x)\,dx\right| < \epsilon(b-a)$,

$$\lim_{h=0} \frac{u_{y_0+h} - u_{y_0}}{h} = \int_a^b \left\{\frac{\partial f(x, y)}{\partial y}\right\}_{y=y_0} dx;$$

and the limit on the left-hand side of this equation is $\left(\frac{du}{dy}\right)_{y=y_0}$, in case u has a differential coefficient at y_0; otherwise this limit is the derivative of u at y_0 on the right, which has thus been shewn to exist, subject to the conditions assumed.

The following theorem has thus been established:—

If $f(x, y)$ be defined in the domain by $a \leqq x \leqq b$, $y_0 \leqq y \leqq y_0 + \alpha$, and $u = \int_a^b f(x, y)\,dx$ exist as a proper integral, for every value of y in that domain, it is a sufficient condition that u may have a derivative with respect to y, at y_0 on the right, and that this derivative be equal to $\int_a^b \left\{\frac{\partial f(x, y)}{\partial y}\right\}_{y=y_0} dx$, that $\frac{\partial f(x, y)}{\partial y}$ should exist everywhere in the domain, and should converge to $\left\{\frac{\partial f(x, y)}{\partial y}\right\}_{y=y_0}$, uniformly for all values of x in (a, b).

In particular, the condition stated in the theorem is satisfied if $\frac{\partial f(x, y)}{\partial y}$ is continuous with respect to (x, y), for all points such that $a \leqq x \leqq b$, $y_0 \leqq y \leqq y_0 + \alpha$.

It is however not necessary for the validity of the process that $\frac{\partial f(x, y)}{\partial y}$ should exist for values of y which are $> y_0$. We shall assume that $f(x, y)$ has for $y = y_0$, either a differential coefficient with respect to y, or at least a definite derivative $Df(x, y_0)$ at y_0 on the right; and this for every value of x in (a, b).

The function $\frac{f(x, y_0 + h) - f(x, y_0)}{h}$ is a function of h, of which the limit for $h = 0$, is $Df(x, y_0)$; and we therefore apply the condition in one of the theorems of § 383, to obtain a sufficient condition that $\int_a^b Df(x, y_0)\,dx$ may exist and be equal to Du. We thus obtain the following theorem:—

If $f(x, y)$ be defined in the domain for which $a \leqq x \leqq b$, $y_0 \leqq y \leqq y_0 + \alpha$, and have a proper integral in (a, b), for each value of y in the interval $(y_0, y_0 + \alpha)$, it is a sufficient condition that $\int_a^b f(x, y)\,dx$ may have a definite*

* This theorem was given, with a direct proof, by G. H. Hardy, *Messenger of Math.*, vol. XXXI, p. 133.

derivative on the right with respect to y at y_0, and that this derivative may be equal to $\int_a^b Df(x, y_0)\,dx$, where D denotes differentiation with respect to y on the right at y_0, that $\frac{f(x, y_0+h)-f(x, y_0)}{h}$ should converge to $Df(x, y_0)$ uniformly for all values of x in (a, b).

It is here not assumed that $Df(x, y)$ has a definite value for any value of y except y_0.

By applying the more general result in § 383, we obtain the following theorem:—

If $f(x, y)$ have a proper integral in the interval (a, b), for each value of y, such that $y_0 \leqq y \leqq y_0 + \alpha$, and $\int_a^b Df(x, y_0)\,dx$ exist as a proper integral, it is a sufficient condition that $\int_a^b f(x, y)\,dx$ may have a definite derivative at y_0 on the right, equal to $\int_a^b Df(x, y_0)\,dx$, that $\left|\frac{f(x, y_0+h)-f(x, y_0)}{h}\right|$ should be less than some fixed number, for all values of x in (a, b), and for all values of h which are such that $0 < h \leqq \alpha$.

For a fixed value of x, the upper limit of $\left|\frac{f(x, y_0+h)-f(x, y_0)}{h}\right|$ cannot exceed the upper limit of the four derivatives $D^+f(x, y)$, $D_+f(x, y)$, $D^-f(x, y)$, $D_-f(x, y)$ in the interval $(y_0, y_0+\alpha)$ of y. It follows that *the second condition contained in the above theorem is satisfied if the derivatives of $f(x, y)$ with respect to y, whether definite or not, are limited in the whole domain for which $a \leqq x \leqq b$, $y_0 \leqq y \leqq y_0 + \alpha$.*

416. If we no longer assume that $\int_a^b f(x, y)\,dx$ exists as a proper integral, for all values of y in the interval (y_0, y_0+h), but may for some, or all such values of y, be an improper integral, we may apply the theorem of § 386, to obtain sufficient conditions for the validity of the rule of differentiation under the integral sign.

We thus obtain the following theorem:—

If the derivative $Df(x, y_0)$ at y_0 on the right exist as a definite finite number, for all points x in (a, b) which do not belong to a reducible closed set of points G, then, under the conditions (1) that in any interval (α, β) contained in (a, b) which is free in its interior and at its ends from points of G, $\left|\frac{f(x, y_0+h)-f(x, y_0)}{h}\right|$ is less than some fixed finite number, for all values of x in (α, β), and for all values of h, such that $0 < h \leqq \alpha$; (2) that $\int_a^b f(x, y)\,dx$ exists at least as an improper integral, for all values of y such *that* $y_0 < y \leqq y_0 + \alpha$; *and* (3) *that* $\int_a^x f(x, y)\,dx$ *has a definite derivative on the right at* y_0, *which is a continuous function of* x *in the interval* (a, b) *of* x; *and* (4) *that* $\int_a^b Df(x, y_0)\,dx$ *exists, at least as an improper integral; the equality of* $\int_a^b Df(x, y_0)\,dx$ *with the derivative of* $\int_a^b f(x, y)\,dx$ *at* y_0 *on the right holds.*

417. The validity of the application of the rule for the differentiation of $\int_a^b f(x, y)\,dx$ with respect to y is, under certain restrictions, dependent on the equality of two repeated integrals. Let us assume that $Df(x, y)$ has a proper integral with respect to y in the interval $(y_0, y_0 + \alpha)$, for each value of x in (a, b); we have then

$$f(x, y_0 + h) - f(x, y_0) = \int_{y_0}^{y_0+h} Df(x, y)\,dy.$$

Since now

$$\frac{u_{y_0+h} - u_{y_0}}{h} = \int_a^b \frac{f(x, y_0 + h) - f(x, y_0)}{h}\,dx$$

$$= \frac{1}{h}\int_a^b dx \int_{y_0}^{y_0+h} Df(x, y)\,dy;$$

provided the function $Df(x, y)$ has a proper integral with respect to (x, y) in the domain bounded by $x = a$, $x = b$, $y = y_0$, $y = y_0 + \alpha$, the order of integration in the repeated integral may be reversed, and we find

$$\frac{u_{y_0+h} - u_{y_0}}{h} = \frac{1}{h}\int_{y_0}^{y_0+h} dy \int_a^b Df(x, y)\,dx.$$

If now $\int_a^b Df(x, y)\,dx$ be continuous with respect to y at the point y_0, the limit of the expression on the right-hand side of this equation is, for $h = 0$, $\int_a^b Df(x, y_0)\,dx$. The following theorem has therefore been established :—

It is sufficient for the validity of the rule of differentiation that $Df(x, y)$ *should have a proper integral with respect to* y, *in the domain bounded by* $x = a$, $x = b$, $y = y_0$, $y = y_0 + \alpha$, *and should also have a proper double integral with respect to* (x, y) *in that domain, and further that* $\int_a^b Df(x, y)\,dx$ *should be continuous with respect to* y, *for* $y = y_0$.

It has here not been assumed that $Df(x, y)$ has a definite value for every value of y such that $y_0 < y \leqq y_0 + \alpha$, but only that it is integrable in the interval $(y_0, y_0 + \alpha)$ of y.

418. In case the upper limit b have the improper value ∞, the condition that $Df(x, y)$ shall have an integral in the domain bounded by $y = y_0$, $y = y_0 + \alpha$, and unbounded in the positive direction of x is not sufficient to ensure that the repeated integral may be reversed. It may in fact happen that $\int_a^\infty Df(x, y)\,dy$ does not exist. If however we assume that the conditions stated in the theorem of § 409 are satisfied, the process is still valid, and the rule of differentiation is still applicable.

In this case, $b = \infty$, we have

$$u_{y_0+h} - u_{y_0} = \lim_{X=\infty} \int_a^X \{f(x, y_0 + h) - f(x, y_0)\}\,dx$$
$$= \lim_{X=\infty} \int_a^X dx \int_{y_0}^{y_0+h} Df(x, y)\,dy,$$

it being assumed as before that, in the domain bounded by $x = a$, $x = X$, $y = y_0$, $y = y_0 + \alpha$, Df exists for all values of y, and is integrable with respect to y; and that this holds however great X may be. On the assumption that $Df(x, y)$ has a proper integral in the same domain, we have

$$u_{y_0+h} - u_{y_0} = \lim_{X=\infty} \int_{y_0}^{y_0+h} dy \int_a^X Df(x, y)\,dx. \qquad \text{(A)}$$

If now

$$\lim_{X=\infty} \int_{y_0}^{y_0+h} dy \int_X^\infty Df(x, y)\,dy = 0, \qquad \text{(B)}$$

and further if $\int_a^\infty Df(x, y)\,dx$ be continuous with respect to y at $y = y_0$, we have

$$\left(\frac{\partial u}{\partial y}\right)_{y=y_0} = \int_a^\infty Df(x, y)\,dx.$$

In case $\int_a^\infty Df(x, y)\,dx$ does not exist, or the equation (B) be otherwise not valid, the equation (A) still holds, and it may in certain cases be applied to determine the value of $\left(\frac{\partial u}{\partial y}\right)_{y=y_0}$.

Let us assume* that $\int_a^X Df(x, y)\,dx$ can be divided into two components, so that

$$\int_a^X Df(x, y)\,dx = \phi(X, y) + \int_a^X \psi(x, y)\,dx$$

where $\phi(X, y)$ is such that $\lim_{X=\infty} \int_{y_0}^{y_0+h} \phi(X, y)\,dy = 0$, and where $\psi(x, y)$ is such that

$$\int_a^\infty dx \int_{y_0}^{y_0+h} \psi(x, y)\,dy = \int_{y_0}^{y_0+h} dy \int_a^\infty \psi(x, y)\,dx.$$

* De la Vallée-Poussin, *Annales de la soc. scien. de Bruxelles*, vol. XVI, 2.

We find then, provided $\int_a^\infty \psi(x, y)\,dx$ is a continuous function of y at the point $y = y_0$, that

$$\left(\frac{\partial u}{\partial y}\right)_{y=y_0} = \int_a^\infty \psi(x, y_0)\,dx.$$

419. In ordinary cases, a special case of the criterion of § 409 may be applied to prove the validity of the inversion involved in the use of the equation

$$\int_a^\infty dx \int_{y_0}^{y_0+h} Df(x, y)\,dy = \int_{y_0}^{y_0+h} dy \int_a^\infty Df(x, y)\,dx;$$

and then, provided $\int_a^\infty Df(x, y)\,dx$ is continuous with respect to y for $y = y_0$, we have

$$\left(\frac{\partial u}{\partial y}\right)_{y=y_0} = \int_a^\infty \{Df(x, y)\}_{y=y_0}\,dx.$$

It is thus established that *a sufficient condition for the differentiability of $\int_a^\infty f(x, y)\,dx$ at y_0, under the sign of integration, is that $\int_a^\infty Df(x, y)\,dx$ shall converge uniformly for all values of y in the interval $(y_0, y_0 + \alpha)$, and shall be a continuous function of y at $y = y_0$.*

It may be observed that *the condition that $\int_a^\infty Df(x, y)\,dx$ shall be a continuous function of y at y_0, may be replaced by the condition that $\int_a^X Df(x, y)\,dx$ be continuous, whatever value X may have $(> a)$, it being assumed that the condition of uniform convergence of $\int_a^\infty Df(x, y)\,dx$ is satisfied.*

For
$$\int_a^\infty Df(x, y)\,dx = \int_a^X Df(x, y)\,dx + \eta(y)$$
where $|\eta(y)| < \epsilon$, provided X is sufficiently great.

Hence

$$\int_a^\infty Df(x, y_0 + h)\,dx - \int_a^\infty Df(x, y_0)\,dx$$
$$= \int_a^X Df(x, y_0 + h)\,dx - \int_a^X Df(x, y_0)\,dx + \zeta,$$

where $|\zeta| < 2\epsilon$. From this it follows that, for a sufficiently small value of h,

$$\int_a^\infty Df(x, y_0 + h)\,dx - \int_a^\infty Df(x, y_0)\,dx$$

is less than 3ϵ; and since ϵ is arbitrarily small, $\int_a^\infty Df(x, y)\,dx$ is continuous at $y = y_0$.

420. The method employed in § 417 may be extended to a wider class of cases. It may happen that $Df(x, y)$, although limited in the domain bounded by $x=a$, $x=b$, $y=y_0$, $y=y_0+\alpha$, is not integrable with respect to y in the interval (y_0, y_0+h), for all values of x. In that case, we can employ the Lebesgue integral $\int_{y_0}^{y_0+h} Df(x, y)\,dy$, which certainly exists for every value of y, since, $f(x, y)$ being summable, $Df(x, y)$ is also summable, provided it have at each point a definite value, with the possible exception of a set of points of zero measure. In accordance with the theorem of § 406, the order of the repeated integrals can be reversed, provided

$$\int_{y_0}^{y_0+h} dy \int_a^b Df(x, y)\,dx$$

have a definite meaning; and this is certainly the case, since

$$\left|\int_a^b Df(x, y)\,dx\right|$$

cannot exceed the product of $b-a$ into the upper limit of $|Df(x, y)|$ in the two-dimensional domain. The reasoning of § 417 is then applicable, and we obtain the following theorem:—

If $Df(x, y)$ be limited and in general definite, in the domain bounded by $x=a$, $x=b$, $y=y_0$, $y=y_0+\alpha$, then, provided the Lebesgue integral

$$\int_a^b Df(x, y)\,dx$$

be continuous at y_0, on the right, the derivative of

$$\int_a^b f(x, y)\,dx,$$

at y_0 on the right, is $$\int_a^b Df(x, y_0)\,dx.$$

It may happen that $Df(x, y)$ is unlimited in the domain bounded by

$$x=a,\quad x=b,\quad y=y_0,\quad y=y_0+\alpha,$$

but may still possess a Lebesgue integral.

Also $\int_{y_0}^{y_0+h} Df(x, y)\,dy$ may exist as a Lebesgue integral, although $Df(x, y)$ is not necessarily limited in the interval (y_0, y_0+h). In case the points of infinite discontinuity form a reducible set of points in the interval (y_0, y_0+h), the equation

$$f(x, y_0+h)-f(x, y_0)=\int_{y_0}^{y_0+h} Df(x, y)\,dy$$

is still valid. If further, the repeated integral

$$\int_{y_0}^{y_0+h} dy \int_a^b Df(x, y)\,dx$$

have a definite meaning, as a repeated Lebesgue integral, then the process of § 417 is still valid. We have accordingly obtained the following theorem, applicable to those cases in which $Df(x, y)$ is not necessarily a limited function, but may have points of infinite discontinuity :—

If the points of infinite discontinuity of $Df(x, y)$, considered as a function of y, form a reducible set, for each value of x, and

$$\int_{y_0}^{y_0+h} Df(x, y)\,dy$$

exist as a Lebesgue integral, for each value of x, and if $Df(x, y)$ have a Lebesgue double integral in the domain bounded by

$$x = a,\quad x = b,\quad y = y_0,\quad y = y_0 + \alpha,$$

and if $\int_{y_0}^{y_0+h} dy \int_a^b Df(x, y)\,dx$ exist as a repeated Lebesgue integral, and

$$\int_a^b Df(x, y)\,dx$$

be continuous at $y = y_0$, then the derivative of $\int_a^b f(x, y)\,dx$ at y_0 on the right, is $\int_a^b Df(x, y_0)\,dx$.

In any particular case the Lebesgue integrals may exist in accordance with the older definitions.

EXAMPLES.

1*. Let

$$f(x, y) = \sin\left(4 \tan^{-1}\frac{y}{x}\right) - \frac{4xy}{x^2+y^2}\cos\left(4 \tan^{-1}\frac{y}{x}\right);$$

then

$$\int_0^X f(x, y)\,dx = X \sin\left(4 \tan^{-1}\frac{y}{X}\right).$$

We find

$$\frac{\partial}{\partial y}\int_0^X f(x, y)\,dx = \frac{4X^2}{X^2+y^2}\cos\left(4 \tan^{-1}\frac{y}{X}\right);$$

therefore at the point $y=0$,

$$\frac{\partial}{\partial y}\int_0^X f(x, y)\,dx = 4.$$

The value of $\int_0^X \frac{\partial f(x, y)}{\partial y}\,dx$ is found to be $\frac{X^2}{X^2+y^2}\cos\left(4\tan^{-1}\frac{y}{x}\right)$, when $y>0$, and zero when $y=0$. Since this integral is not continuous at $y=0$, the conditions for the differentiation under the sign of integration are not satisfied at $y=0$; in fact we have

$$\int_0^X \frac{\partial f(x, 0)}{\partial y}\,dx = 0.$$

The function $f(x, y)$ is discontinuous at $x=0$, $y=0$.

* Harnack's *Diff. and Int. Calc.*, Cathcart's translation, p. 266.

2. Consider the integral $\int_0^\infty \frac{\sin xy}{x}\,dx$, where $y>0$. This integral is not differentiable under the sign of integration, for any value of y; for $\int_0^\infty \cos xy\,dx$ does not exist.

3*. The integral $\int_0^X (x-y)^{\frac{1}{3}}\,dx$ may be differentiated under the sign of integration, for every value of y. For it has been shewn in Ex. 3, § 408, that $(x-y)^{-\frac{2}{3}}$ has an absolutely convergent double integral in the domain bounded by $x=0$, $x=X$, $y=0$, $y=h$. It may be easily verified that the other conditions of the second theorem in § 420 are satisfied.

4. Consider the integral $u=\int_0^\infty \frac{\cos xy}{1+x^2}\,dx$, where $y>0$. The integral

$$\int_0^\infty \frac{x\sin xy}{1+x^2}\,dx$$

satisfies the condition that, for all values of y greater than a positive number y_0, it converges uniformly. We find, by integration by parts, that

$$\int_X^{X'} \frac{x\sin xy}{1+x^2}\,dx=\left[-\frac{x\cos xy}{(1+x^2)\,y}\right]_X^{X'}+\frac{1}{y}\int_X^{X'}\frac{1-x^2}{(1+x^2)^2}\cos xy\,dx\,;$$

hence if $X'>X>1$, the absolute value of the integral is less than

$$\frac{2X}{(1+X^2)\,y_0}+\frac{1}{y_0}\left(\frac{\pi}{2}-\tan^{-1}X\right);$$

whence the result follows. The condition for differentiation of u under the integral sign is therefore satisfied, for any positive value of y.

THE CONDENSATION OF SINGULARITIES.

421. A method of constructing functions which possess, at an infinitely numerous set of points in a linear interval, singularities in relation to continuity, derivatives, or oscillations, has been given by Hankel. The method depends upon the employment of functions which at a single point possess one of the singularities in question, and consists in building up by the use of such a function of a simple type, the more complicated analytical representations of a function which possess the singularity at an everywhere-dense set of points. To this method, Hankel† has given the name "Principle of condensation of singularities" (das Prinzip der Verdichtung der Singularitäten); the name may however be conveniently applied to other methods of constructing functions capable of analytical representation which have been given more recently by other writers.

Let $\phi(y)$ be a function defined for the interval $(-1,+1)$, limited in that interval, and continuous at every point of the interval, including $-1, +1$,

* See Hardy, *Quarterly Journal Math.*, vol. XXXII, p. 67, where various theorems relating to the differentiation of integrals will be found.

† See his memoir "Untersuchungen über die unendlich oft unstetigen im oscillierenden Functionen," reproduced in *Math. Annalen*, vol. XX. The method has been treated in a rigorous manner by Dini, *Grundlagen*, p. 157.

except at the point $y=0$, where however $\phi(0)=0$. The function $\phi(\sin n\pi x)$ is finite and continuous for every value of x which is not a rational fraction m/n, with n as denominator, and it vanishes for all points at which x has this form.

The series $\sum_{n=1}^{\infty}\frac{\phi(\sin n\pi x)}{n^s}$, where $s>1$, is, in accordance with the fact that $\phi(y)$ is limited, uniformly convergent in every interval; and its sum is a limited function of x which is continuous for all irrational values of x.

If $\phi(y)$ were also continuous for $y=0$, the function represented by the series would be continuous also for rational values of x, but when $\phi(y)$ is discontinuous at $y=0$, the properties of the function

$$f(x)\equiv\sum_{n=1}^{\infty}\frac{\phi(\sin n\pi x)}{n^s}$$

in relation to continuity or discontinuity at the points where x has rational values require investigation.

The series being uniformly continuous, it follows from the theorem of § 343, that $f(x)$ is continuous at every point at which all the functions $\phi(\sin n\pi x)$ are continuous, *i.e.* for all irrational values of x. Let us consider the values of the function $f(x)$ in the neighbourhood of a point $x=p/q$, where p and q are relative primes. We may write the value of $f(x)$ in the form

$$\sum_{n_q=1}^{\infty}\frac{\phi(\sin n_q\pi x)}{n_q^{\,s}}+\frac{1}{q^s}\sum_{m=1}^{\infty}\frac{\phi(\sin qm\pi x)}{m^s},$$

where n_q has those integral values only which are not multiples of q.

The first of these series is uniformly convergent, and its sum is continuous at the point p/q; we therefore find that

$$f(p/q+h)-f(p/q)=\eta_h+\frac{1}{q^s}\sum_{m=1}^{\infty}\frac{\phi(\overline{-1}|^{mp}\sin qm\pi h)}{m^s},$$

where η_h converges to zero when h does so.

Case I. Let $\phi(y)$ have an ordinary discontinuity at $y=0$; we then have

$$f(p/q+0)-f(p/q)=\frac{\phi(\pm 0)}{q^s}\sum_{r=0}^{\infty}\frac{1}{(2r+1)^s}+\frac{\phi(0)}{q^s}\sum_{r=1}^{\infty}\frac{1}{(2r)^s},$$

$$f(p/q-0)-f(p/q)=\frac{\phi(\mp 0)}{q^s}\sum_{r=0}^{\infty}\frac{1}{(2r+1)^s}+\frac{\phi(-0)}{q^s}\sum_{r=1}^{\infty}\frac{1}{(2r)^s},$$

where the upper or lower of the ambiguous signs are to be taken, according as p is even or odd.

If $\phi(+0)$, $\phi(-0)$ are different from one another, and from zero, these relations shew that, at a point p/q for which p is even, the function $f(x)$ has ordinary discontinuities both on the right and on the left, the measures of the two being not identical. Moreover the same statement may be made for a point

p/q at which p is odd, unless $\phi(+0)$, $\phi(-0)$ have such values that one or other of the above expressions vanishes, in which case there is an ordinary discontinuity on one side, and the function is continuous on the other side. It is easily seen to be impossible that the two expressions can simultaneously vanish, and therefore there is an ordinary discontinuity on one side at least.

If $\phi(+0) \neq 0$, $\phi(-0) = 0$, there is discontinuity on the right at the points $x = 2p'/q$, and continuity on the left; and at the points $x = (2p'+1)/q$, there are discontinuities on both sides, with different measures.

If $\phi(+0) = \phi(-0)$, so that $\phi(y)$ has only a removable discontinuity at the point $y = 0$, then the function $f(x)$ has removable discontinuities at all the points $x = p/q$.

In every case the function $f(x)$ is a point-wise discontinuous function, because its discontinuities are all ordinary ones (see § 189).

Case II. Let $\phi(y)$ have a discontinuity of the second kind, at $y = 0$, on one side at least. In this case it will be assumed that $s > 2$. Denoting by A the upper limit of $|\phi(y)|$ in the interval $(-1, +1)$, we have

$$\left| \sum_{m=1}^{\infty} \frac{\phi(\overline{-1}|^{mp} \sin qm\pi h)}{m^s} - \phi(\overline{-1}|^{p} \sin q\pi h) \right|$$
$$< \frac{A}{2^{s-2}} \sum_{m=1}^{\infty} \frac{1}{(m+1)^2} < \frac{A}{2^{s-2}} \left(\frac{\pi^2}{6} - 1\right),$$

and hence

$$f(p/q+h) - f(p/q) = \frac{1}{q^s} \phi(\overline{-1}|^{p} \sin q\pi h) + \eta_h + \frac{\zeta A}{2^{s-2}} \cdot \frac{1}{q^s},$$

where ζ is such that $-1 < \zeta < 1$, and is dependent on h.

If $\phi(y)$ have a discontinuity of the second kind on both sides of the point $y = 0$, there are finite oscillations in arbitrarily small neighbourhoods of the point on the two sides; if then s be chosen so great that $A/2^{s-2}$ is less than half the saltus at $y = 0$, we see that $f(x)$ has discontinuities of the second kind on both sides at all the points $x = p/q$.

If $\phi(y)$ have a discontinuity of the second kind at $y = 0$, on the right, and have a discontinuity of the first kind, or be continuous, on the left, there is at each of the points $x = p/q$, where p is even, a discontinuity of $f(x)$ of exactly the same kind as that of $\phi(y)$ at $y = 0$. On the other hand, if s be sufficiently large, there is at each of the points $x = p/q$, where p is odd, a discontinuity of the second kind on both sides. For we may express $f(p/q+h) - f(p/q)$ in the form

$$\eta_h + \frac{1}{q^s} \sum_{r=0}^{\infty} \frac{\phi(-\sin \overline{2r+1}\, q\pi h)}{(2r+1)^s} + \frac{1}{q^s} \sum_{r=0}^{\infty} \frac{\phi(\sin 2rq\pi h)}{(2r)^s},$$

which can, as in the previous case, be reduced to the form

$$\eta_h + \frac{1}{q^s}\phi(-\sin q\pi h) + \frac{1}{q^s}\phi(\sin 2q\pi h) + \frac{A\zeta_1}{q^s 3^{s-2}} + \frac{A\zeta_2}{q^s 2^{s-2}},$$

where ζ_1, ζ_2 are both in the interval $(-1, 1)$. We thus see that, if s be sufficiently great, there are finite oscillations in arbitrarily small neighbourhoods of p/q on both sides.

The existence of the factor $1/q^s$ in the expression for $f(p/q+h)-f(p/q)$ shews that there are only a finite number of points p/q at which the saltus of $f(x)$ is $\geqq k$, where k is an arbitrarily chosen positive number; and thus $f(x)$ belongs to the special class of point-wise discontinuous functions for which the set K is a finite set, for each value of k.

EXAMPLES.

1. Let $\phi(y) = \sin\frac{1}{y}$, and $\phi(0)=0$; the function $f(x)$ is then defined by

$$f(x) = \sum_1^\infty \frac{1}{n^s}\sin(\operatorname{cosec} n\pi x),$$

where, when $x=p/q$, the terms for which n is a multiple of q are to be omitted. This function is, at least when $s>2$, a point-wise discontinuous function which is continuous at all the irrational points, and has discontinuities of the second kind, on both sides, at the rational points.

2. Let
$$\phi(y) = \frac{4}{\pi}\sum_{r=0}^\infty \frac{1}{(2r+1)}\sin(2r+1)\frac{\pi y}{a},$$

where $a>1$. For $0<y\leqq 1$, we have $\phi(y)=1$; for $-1\leqq y<0$, we have $\phi(y)=-1$; also $\phi(0)=0$; and thus $\phi(y)$ has an ordinary discontinuity at $y=0$.

The function
$$f(x) = \frac{4}{\pi}\sum_{n=1}^\infty \frac{1}{n^s}\left[\sum_{r=0}^\infty \frac{1}{2r+1}\sin\left\{(2r+1)\frac{\pi}{a}\sin n\pi x\right\}\right],$$

is a point-wise discontinuous function, which is continuous for all irrational values of x, and has ordinary discontinuities on both sides at all the rational points.

3. With the same value of $\phi(y)$ as in Ex. (2), let

$$\chi(x) = \sum_1^\infty \frac{1}{n^s[\phi(\sin n\pi x)]^s},$$

where $s>1$. For any irrational value of x, $\chi(x)$ has the value $\sum_1^\infty \frac{1}{n^s}$, and for any rational value of x, the function is indefinitely great. Now let

$$f(x) = \frac{\sum_1^\infty \frac{1}{n^s}}{\chi(x)};$$

then $f(x)=1$, for all irrational values of x, and $f(x)=0$, for all rational values of x. The function $f(x)$ is accordingly totally discontinuous. The values of $f(x)$ are improperly defined at the rational points.

422. Let us next assume that $\phi(y)$ is continuous throughout the interval $(-1, 1)$, and has no differential coefficient at the point $y=0$, where $\phi(0)=0$, but that, at every other point in the interval $(-1, +1)$, it has a differential coefficient which is numerically less than some fixed finite number A. In this case $\frac{\phi(h)}{h}$ has no definite limit for $h=0$, either when h is positive, or when h is negative, or in both cases; or else the two limits both exist, but have different values.

The numbers $\left|\frac{\phi(h)}{h}\right|$ which are equal to $|\phi'(\theta h)|$, where $\theta > 0$, have a definite upper limit $U (\leqq A)$ for all values of h.

Assuming that $s > 2$, we see that the series

$$\pi \sum_1^\infty \frac{\phi'(\sin n\pi x)}{n^{s-1}} \cos n\pi x$$

converges for all irrational values of x, since the general term is numerically less than B/n^{s-1}, where B is some fixed number.

Consider the series

$$\sum_1^\infty \frac{\phi[\sin n\pi(x+h)] - \phi(\sin n\pi x)}{hn^s},$$

where x has an irrational value. It will be shewn that this series converges uniformly for all values of h. Unless n and h are such that $\sin n\pi(x+h)$ and $\sin n\pi x$ are equal, in which case $\phi[\sin n\pi(x+h)] - \phi(\sin n\pi x) = 0$, we can write the general term of the series in the form

$$\frac{\pi}{n^{s-1}} \cdot \frac{\phi[\sin n\pi(x+h)] - \phi(\sin n\pi x)}{\sin n\pi(x+h) - \sin n\pi x} \cdot \frac{\sin \frac{1}{2} n\pi h}{\frac{1}{2} n\pi h} \cdot \cos n\pi(x + \tfrac{1}{2}h).$$

It then follows that the general term of the series is numerically less than $\frac{\pi V}{n^{s-1}}$, where V is the upper limit of the absolute values of the incrementary ratios of the function. Since the series $\Sigma 1/n^{s-1}$ is convergent, it follows that the above series converges uniformly for all values of h which are $\neq 0$; and consequently, in accordance with the last theorem of § 397, the function $f(x)$ has a finite differential coefficient for any irrational value of x.

Next let x have the rational value p/q. We may then express $\frac{f\left(\frac{p}{q}+h\right) - f\left(\frac{p}{q}\right)}{h}$ in the form

$$\sum_{n_q=1}^\infty \frac{\phi[\sin n_q\pi(x+h)] - \phi(\sin n_q \pi x)}{hn_q^s} + \frac{1}{q^s} \sum_{m=1}^\infty \frac{\phi(\overline{-1}|^{mp} \sin mq\pi h)}{hm^s},$$

where n_q has all positive integral values which are not multiples of q. In accordance with the above proof we see that $\sum_{n_q=1}^{\infty} \frac{\phi(\sin n_q \pi x)}{n_q^s}$ has, for the value $x = p/q$, a finite differential coefficient which is the sum

$$\pi \sum_{n_q=1}^{\infty} \frac{\phi'(\sin n_q \pi p/q)}{n_q^{s-1}} \cos n_q \pi p/q.$$

We have now shewn that

$$\frac{f\left(\frac{p}{q}+h\right)-f\left(\frac{p}{q}\right)}{h}$$

$$= \pi \sum_{n_q=1}^{\infty} \frac{\phi'(\sin n_q \pi p/q)}{n_q^{s-1}} \cos n_q \pi p/q + \eta_h + \frac{1}{q^s} \sum_{m=1}^{\infty} \frac{\phi(\overline{-1}|^{mp} \sin mq\pi h)}{hm^s},$$

where η_h is a number which converges to zero when h is indefinitely diminished.

Case I. Let $\phi(y)$ have definite derivatives on the right and on the left when $y=0$; and thus $\frac{\phi(h)}{h}$ has one limit $\phi'(+0)$ for positive values of h converging to zero, and another limit $\phi'(-0)$ for negative values of h so converging. We thus have, when p is even,

$$\lim_{h=+0} \frac{f\left(\frac{p}{q}+h\right)-f\left(\frac{p}{q}\right)}{h} = \pi \sum_{n_q=1}^{\infty} \frac{\phi'(\sin n_q \pi p/q)}{n_q^{s-1}} \cos n_q \pi p/q + \frac{\pi\phi'(+0)}{q^{s-1}} \sum_{m=1}^{\infty} \frac{1}{m^{s-1}},$$

$$\lim_{h=-0} \frac{f\left(\frac{p}{q}+h\right)-f\left(\frac{p}{q}\right)}{h} = \pi \sum_{n_q=1}^{\infty} \frac{\phi'(\sin n_q \pi p/q)}{n_q^{s-1}} \cos n_q \pi p/q + \frac{\pi\phi'(-0)}{q^{s-1}} \sum_{m=1}^{\infty} \frac{1}{m^{s-1}}.$$

For an uneven value of p, we find

$$\lim_{h=+0} \frac{f\left(\frac{p}{q}+h\right)-f\left(\frac{p}{q}\right)}{h} = \pi \sum_{n_q=1}^{\infty} \frac{\phi'(\sin n_q \pi p/q)}{n_q^{s-1}} \cos n_q \pi p/q + \frac{\pi\phi'(+0)}{q^{s-1}} \sum_{r=1}^{\infty} \frac{1}{(2r)^{s-1}} - \frac{\pi\phi'(-0)}{q^{s-1}} \sum_{r=0}^{\infty} \frac{1}{(2r+1)^{s-1}},$$

$$\lim_{h=-0} \frac{f\left(\frac{p}{q}+h\right)-f\left(\frac{p}{q}\right)}{h} = \pi \sum_{n_q=1}^{\infty} \frac{\phi'(\sin n_q \pi p/q)}{n_q^{s-1}} \cos n_q \pi p/q - \frac{\pi\phi'(+0)}{q^{s-1}} \sum_{r=1}^{\infty} \frac{1}{(2r+1)^{s-1}} + \frac{\pi\phi'(-0)}{q^{s-1}} \sum_{r=1}^{\infty} \frac{1}{(2r)^{s-1}}.$$

From these results it is seen that $f(x)$ has, at the rational points, definite derivatives on the right and on the left, differing in value from one another, and therefore, at all these points, the function has a singularity of the same kind as $\phi(y)$ possesses at the point $y=0$.

Case II. Let $\phi(y)$ have, on one side of $y=0$ at least, no definite derivative. Unless mqh is an integer, in which case $\phi(\overline{-1}|^{mp}\sin mq\pi h)=0$, we have

$$\frac{\phi(\overline{-1}|^{mp}\sin mq\pi h)}{hm^s}=\frac{\phi(\overline{-1}|^{mp}\sin mq\pi h)}{\overline{-1}|^{mp}\sin mq\pi h}\cdot\frac{\sin mq\pi h}{mq\pi h}(-1)^{mp}\frac{q\pi}{m^{s-1}};$$

and this is numerically less than $\dfrac{q\pi U}{m^{s-1}}$. It follows that

$$\frac{1}{q^s}\sum_{m=1}^{\infty}\frac{\phi(\overline{-1}|^{mp}\sin mq\pi h)}{hm^s}=\frac{1}{q^s}\frac{\phi(\overline{-1}|^{p}\sin q\pi h)}{h}+P,$$

where P is numerically less than $\dfrac{\pi U}{q^{s-1}}\sum\limits_{m=2}^{\infty}\dfrac{1}{m^{s-1}}$. By taking a sufficiently large value of s, the number P may be made as small as we please, and therefore

$$\frac{1}{q^s}\sum_{m=1}^{\infty}\frac{\phi(\overline{-1}|^{mp}\sin mq\pi h)}{hm^s}$$

will, for a sufficiently large value of s, oscillate in the same manner as

$$\frac{1}{q^s}\frac{\phi(\overline{-1}|^{p}\sin q\pi h)}{h},$$

as h is diminished indefinitely. It is thus seen that $\dfrac{f\left(\frac{p}{q}+h\right)-f\left(\frac{p}{q}\right)}{h}$ has, on one side, or on both sides, of $h=0$, no definite limit; and thus $f(x)$ has no differential coefficient at any of the rational points, provided a sufficiently large value of s be chosen.

EXAMPLES.

1. Let $\phi(y)=y$ or $-y$, according as y is positive or negative. The corresponding function $f(x)$ is $\sum\limits_1^{\infty}\dfrac{1}{n^s}\sqrt{\sin^2 n\pi x}$, where the positive value of the square root is to be taken. This function is continuous, and has a differential coefficient for all irrational values of x. At the rational points it has no differential coefficient, but has definite derivatives on both sides.

2. Let $\phi(y)=y\sin(\log y^2)$, then $f(x)=\sum\limits_1^{\infty}\dfrac{\sin n\pi x\,[\log\sin^2 n\pi x]}{n^s}$.

The function $f(x)$ is continuous, and has a finite differential coefficient for all irrational values of x. If s be sufficiently large, it has no definite derivatives either on the right or on the left, for rational values of x; the four derivatives at such a point are all finite.

423. Let it next be assumed that $\phi(y)$ is continuous in the interval $(-1, 1)$, and has a finite differential coefficient at every point except at $y=0$, but that this differential coefficient has no upper limit to its absolute magnitude in any neighbourhood of the point $y=0$. In this case $\phi(y)$ may either have a differential coefficient at $y=0$, which is finite or indefinitely great; or it may have indefinitely great derivatives, on the right and on the left, of opposite signs; or it may have no definite derivatives. When $\phi(y)$ is a function of this type, it is not certain that $f'(x)$ has differential coefficients for irrational values of x; for the differential coefficients $\phi'(\sin n\pi x)$ are not all numerically less than a fixed finite number, for such a value of x, and for all values of n; and thus the argument of § 422 does not apply.

For a rational point $x=p/q$, we have as before,

$$\frac{f\left(\frac{p}{q}+h\right)-f\left(\frac{p}{q}\right)}{h}=\sum_{n_q=1}^{\infty}\frac{\phi[\sin n_q\pi(x+h)]-\phi(\sin n_q\pi x)}{hn_q^{s}}$$
$$+\frac{1}{q^s}\sum_{m=1}^{\infty}\frac{\phi(\overline{-1}|^{mp}\sin mq\pi h)}{hm^s}.$$

The theorem of § 398 will be applied to shew that the function $\sum_{n_q=1}^{\infty}\frac{\phi(\sin n_q\pi x)}{n_q^{s}}$ has, for the value p/q of x, a differential coefficient obtained by means of term by term differentiation of the series.

The first condition required by the theorem in question, viz. that the terms of the series

$$\sum_{n_q=1}^{\infty}\frac{\pi\phi'(\sin n_q\pi p/q)}{n_q^{s-1}}\cos n_q\frac{p}{q}\pi$$

shall be definite, and that this series shall converge, is certainly satisfied.

To shew that the conditions relating to

$$\left|\frac{R_m\left(\frac{p}{q}+h\right)}{h}\right|,\quad \left|\frac{R_m\left(\frac{p}{q}\right)}{h}\right|$$

are satisfied, we observe that $R_m(x)<\frac{U}{m^{s-1}}$, where U denotes the upper limit of $|\phi(y)|$ in the interval $(-1, 1)$. Let t be so chosen that $1>t>\frac{1}{s-1}$, $(s>2)$, and let m be that integer next greater than $|h|^{-t}$ which is not a multiple of q; we then have $|hm^{s-1}|>|h|^{1-(s-1)t}$. It follows that, for each fixed value of h, m has been so chosen that

$$\left|\frac{R_m\left(\frac{p}{q}+h\right)}{h}\right|,\quad \left|\frac{R_m\left(\frac{p}{q}\right)}{h}\right|$$

are both less than $U|h|^{(s-1)t-1}$, and are therefore both less than ϵ, provided $|h| < \delta$; where δ is fixed so that $U\delta^{(s-1)t-1} < \epsilon$. It is clear that δ may be chosen so small that m exceeds an arbitrarily prescribed integer m', for all the values of h such that $|h| < \delta$.

We have lastly to prove that the sum of the first m terms of the series of which the general term is

$$\frac{\phi\left[\sin n_q\pi\left(\frac{p}{q}+h\right)\right]-\phi\left(\sin n_q\pi\frac{p}{q}\right)}{hn_q^s}-\frac{\pi\phi'\left(\sin n_q\pi\frac{p}{q}\right)}{n_q^{s-1}}\cos n_q\frac{p}{q}\pi,$$

is numerically less than ϵ.

This series may be divided into two portions $\sum\limits_1^{m_1-1}$ and $\sum\limits_{m_1}^{m}$, where m_1 is a fixed number independent of h, so chosen that the sum $\sum\limits_{m_1}^{\infty}\frac{1}{n_q^{s-1}}$ is less than an arbitrarily chosen number η. The sum of the first m_1 terms of the series under consideration can be made arbitrarily small, by taking δ sufficiently small; for the number of terms is independent of h. We have then only to consider the sum $\sum\limits_{m_1}^{m}$.

Since $n_q\frac{p}{q}$ differs from an integer by at least $\frac{1}{q}$, it follows that $\phi'\left(\sin n_q\pi\frac{p}{q}\right)$ is numerically less than some fixed positive number U', for all values of n_q. We therefore see that

$$\left|\sum_{m_1}^{m}\frac{\pi\phi'\left(\sin n_q\pi\frac{p}{q}\right)}{n_q^{s-1}}\cos n_q\frac{p}{q}\right| < \pi\eta U'.$$

Further, m has been so chosen that $m-|h|^{-t}<1$; from which we have $m|h| = |h|^{1-t}+\theta|h|$, where $0<\theta<1$. If δ be now so chosen that $\delta^{1-t}+\delta<1/2q$, the two numbers $n_q\frac{p}{q}$, $n_q\left(\frac{p}{q}+h\right)$ differ from one another by less than $\frac{1}{2q}$; moreover they are never integers, and contain no integer between them, and they differ from the nearest integer by more than $\frac{1}{2q}$. It follows that, for all values of y between $\sin n_q\pi\frac{p}{q}$ and $\sin n_q\pi\left(\frac{p}{q}+\delta\right)$, where n_q has the values belonging to it in the series $\sum\limits_{m_1}^{m}$, $\phi(y)$ has a differential coefficient numerically less than some fixed number U''.

Writing $\dfrac{\phi\left[\sin n_q\pi\left(\dfrac{p}{q}+h\right)\right]-\phi\left(\sin n_q\pi\dfrac{p}{q}\right)}{hn_q^s}$ in the form

$$\frac{\pi}{n_q^{s-1}}\frac{\phi\left[\sin n_q\pi\left(\dfrac{p}{q}+h\right)\right]-\phi\left(\sin n_q\pi\dfrac{p}{q}\right)}{\sin n_q\pi\left(\dfrac{p}{q}+h\right)-\sin\left(n_q\pi\dfrac{p}{q}\right)}\cos n_q\pi\left(\frac{p}{q}+\frac{1}{2}h\right)\frac{\sin\frac{1}{2}n_q\pi h}{\frac{1}{2}n_q\pi h};$$

we see that this term is numerically less than $\dfrac{\pi}{n_q^{s-1}}U''$. It now follows that

$$\left|\sum_{m}^{m_1}\right|<\pi\eta(U''+U');$$

and this is numerically as small as we please, if we choose η and δ sufficiently small. It is therefore possible to choose δ so small that the last of the requisite conditions is satisfied, for all values of $|h|<\delta$.

It has now been proved that

$$\frac{f\left(\dfrac{p}{q}+h\right)-f\left(\dfrac{p}{q}\right)}{h}=\pi\sum_1^\infty\frac{\phi'\left(\sin n_q\pi\dfrac{p}{q}\right)}{n_q^{s-1}}\cos n_q\pi\frac{p}{q}+\sigma$$
$$+\frac{1}{q^s}\sum_{m=1}^{\infty}\frac{\phi(\overline{-1}|^{mp}\sin mq\pi h)}{hm^s},$$

where σ and h converge together to zero.

The second series on the right-hand side of this equation can be written in the form

$$\frac{1}{q^s}\sum_{n=1}^{n=m}\frac{\phi(\overline{-1}|^{np}\sin nq\pi h)}{hn^s}+\frac{1}{q^s}\sum_{n=m+1}^{\infty}\frac{\phi(\overline{-1}|^{np}\sin nq\pi h)}{hn^s},$$

where m is fixed as before, for each value of h. The second sum is arbitrarily small for a sufficiently small value of δ. We have then to consider the first sum, which may be written in the form

$$\frac{\pi}{q^{s-1}}\sum_{n=1}^{n=m}\frac{\overline{-1}|^{np}}{n^{s-1}}\frac{\phi(\overline{-1}|^{np}\sin nq\pi h)}{\overline{-1}|^{np}\sin nq\pi h}\frac{\sin nq\pi h}{nq\pi h};$$

and we now consider this sum in the different cases which arise when various assumptions are made as to the nature of the derivatives of $\phi(y)$ at the point $y=0$.

Case I. Let $\phi(y)$ have the derivative $+\infty$, at $y=0$ on the right, and the derivative $-\infty$, at $y=0$ on the left. It is clear that, for positive values of h, all the terms of the series have one and the same sign, δ having been chosen so small that mh is also sufficiently small; also it is clear that the first term of the series becomes numerically arbitrarily great for sufficiently small values of h. It therefore appears that the sum of the series becomes indefinitely great, as h approaches the limit zero from the right-hand side. If h be negative, the terms of the series all have the same sign, the opposite one from that which they have when h is positive, and as before, the sum of the series is indefinitely great as h converges to zero.

It has therefore been shewn that

$$\frac{f\left(\frac{p}{q}+h\right)-f\left(\frac{p}{q}\right)}{h}$$

has the limit $+\infty$ on one side of the point $\frac{p}{q}$, and $-\infty$ on the other side. The singularities of the derivative of $f(x)$ at the rational points have the same peculiarity as that of $\phi(y)$ at the point $y=0$; *i.e.* derivatives on the right and on the left exist, which are infinite, but of opposite signs.

Case II. Let $\phi(y)$ have a differential coefficient at $y=0$, which is either $+\infty$, or $-\infty$.

It is then clear that, in case p be even,

$$\frac{f\left(\frac{p}{q}+h\right)-f\left(\frac{p}{q}\right)}{h}$$

has the same limit $+\infty$, or $-\infty$, as $\frac{\phi(h)}{h}$. If p be odd, the terms of the series under examination have alternate signs, and no conclusion can in general be drawn as to the nature of the derivatives of $f(x)$ at the point $x=\frac{p}{q}$.

Case III. Let $\phi(y)$ have a finite differential coefficient at $y=0$.

In this case, as is easily seen, $f(x)$ has, at the point $\frac{p}{q}$, a definite differential coefficient of which the value is

$$\pi\sum_{m=1}^{\infty}\frac{\phi'\left(\sin n\pi\frac{p}{q}\right)}{n^{s-1}}\cos n\pi\frac{p}{q}.$$

Case IV. Let $\phi(y)$ have finite derivatives at $y=0$ on the right, and on the left, which differ from one another. In this case $f(x)$ has, at each rational point, finite derivatives on the right, and on the left, which differ from one another.

Case V. Let $D^+\phi(0)$, $D_+\phi(0)$, $D^-\phi(0)$, $D_-\phi(0)$ be all finite and different from one another. The function $f(x)$ has then at $\frac{p}{q}$, at least when p is even, the same peculiarity as $\phi(y)$ at $y=0$.

EXAMPLES.

1. Let $\phi(y) = y \sin\frac{1}{y}$, $\phi(0) = 0$. The corresponding function $f(x)$ is given by

$$f(x) = \sum_1^\infty \frac{\sin n\pi x \sin\frac{1}{n\pi x}}{n^s}, \text{ where } s > 2.$$

This function is continuous, but has no definite derivatives at the rational points. No assertion can be made as to the derivatives at the irrational points, because the differential coefficient $\phi'(y)$ has indefinitely great values in every neighbourhood of $y = 0$.

2. Let $\phi(y) = (y^2)^{\frac{\alpha}{\beta}}$, where α, β are positive integers such that $2\alpha < \beta$, and the real positive values of the root are taken. We then have

$$f(x) = \sum_1^\infty \frac{(\sin^2 n\pi x)^{\frac{\alpha}{\beta}}}{n^s}, \text{ where } s > 2.$$

This function is continuous, and has, at all rational points, indefinitely great derivatives on the right, and on the left, of opposite signs. No assertion can be made as to the derivatives at the irrational points.

CANTOR'S METHOD OF CONDENSATION OF SINGULARITIES.

424. A method of constructing a function which exhibits at an everywhere-dense set of points, some singularity, either in relation to continuity, or to its derivatives, has been given by Cantor *. Let $\phi(y)$ denote a function which is continuous for all values of y in the interval $(-1, 1)$, except $y = 0$; and let $\phi(0) = 0$. Let G denote an enumerable set of points $\omega_1, \omega_2, \omega_3, \ldots$, which may be everywhere-dense. The method of condensation consists of the construction of the function

$$f(x) = \sum_{n=1}^\infty c_n \phi(x - \omega_n),$$

where $c_1, c_2, \ldots, c_n, \ldots$ are positive numbers, so chosen that the series $\sum_1^\infty c_n$ is convergent, and that $\sum_{n=1}^\infty c_n \phi(x - \omega_n)$ converges absolutely for each value of x, and uniformly in every interval.

This method has two advantages over that of Hankel. In the first place, the points $\omega_1, \omega_2, \ldots$ do not necessarily consist of the rational points of the interval $(-1, +1)$, but may form any enumerable aggregate. In the second place, for a value ω_n of x, the singularity in question is exhibited by the one term $c_n \phi(x - \omega_n)$ only, of the series which represents $f(x)$; whereas in Hankel's method, the singularity of $\phi(y)$ at $y = 0$ is exhibited, for $x = p/q$, by an indefinitely great number of terms of the series which represents the function formed by condensation.

* *Math. Annalen*, vol. XIX. See also Dini's *Grundlagen*, p. 188.

Let now $\phi(y)$ be discontinuous at $y=0$; then, for any value of x_0 of x, which is not one of the values of G, the terms of the series $\Sigma c_n \phi(x-\omega_n)$ are all continuous; hence, since the series converges uniformly in any interval containing x_0, it follows that $f(x)$ is continuous at x_0. Again, in order to consider the continuity of $f(x)$ at the point ω_n, we may separate the term $c_n\phi(x-\omega_n)$ from the rest of the series. As before, the series which consists of all the terms except the one $c_n\phi(x-\omega_n)$ represents a function which is continuous at $x=\omega_n$, but $c_n\phi(x-\omega_n)$ has at ω_n a discontinuity of the same character as that of $\phi(y)$ at $y=0$. It has therefore been shewn that $f(x)$ is continuous at every point which does not belong to G, but has at every point of G a discontinuity of the same character as that of $\phi(y)$ at the point $y=0$. If $\phi(y)$ have a finite saltus k at $y=0$, the saltus of $c_n\phi(x-\omega_n)$ at ω_n is kc_n. Hence, on account of the convergence of Σc_n, there are only a finite number of points ω_n at which the saltus of $f(x)$ exceeds any fixed positive number. The function $f(x)$ is therefore an integrable function.

Let it next be assumed that $\phi(y)$ is continuous throughout $(-1, 1)$, and possesses a differential coefficient for every value of y except $y=0$; and that the differential coefficients are all numerically less than some fixed positive number B. It then follows that the four derivatives of $\phi(y)$ at $y=0$ are all finite; it also follows that $\left|\frac{\phi(h)}{h}\right|$ is less than some fixed number A, for all values of h which are numerically less than some fixed number δ.

We now see that for any pair of points y_1, y_2 such that $|y_1-y_2|<\delta$, we have $\left|\frac{\phi(y_1)-\phi(y_2)}{y_1-y_2}\right|<$ the greater of the numbers A and B, which may be denoted by C.

If x be not a point of G, the sum

$$\sum_{n=m+1}^{\infty} c_n \left|\frac{\phi(x+h-\omega_n)-\phi(x-\omega_n)}{h}\right| \text{ is } < C \sum_{n=m+1}^{\infty} c_n,$$

provided $|h|<\delta$; hence the series is uniformly convergent for all values of h such that $0<|h|<\delta$, and therefore it represents the value of $f'(x)$.

In case x be a point ω_n of G, we separate from the series which represents $f(x)$, the term $c_n\phi(x-\omega_n)$. It appears then that the remaining part of the series represents a function which has a definite differential coefficient $\lambda(\omega_n)$ at ω_n.

We have therefore

$$\frac{f(\omega_n+h)-f(\omega_n)}{h}=c_n\frac{\phi(h)}{h}+\lambda(\omega_n)+\zeta,$$

where ζ converges to zero when h does so. It thus appears that $f(x)$ has no definite derivatives at $x=\omega_n$, but that it has at that point the same kind of singularity as $\phi(y)$ has at the point $y=0$.

EXAMPLES.

1. Let $\phi(y) = y - \frac{1}{2}y \sin(\frac{1}{2}\log x^2)$. This function has a differential coefficient $\phi'(y)$ for every point except $y=0$; and $\phi'(y)$ oscillates between the values $1 - 1/\sqrt{2}$, $1 + 1/\sqrt{2}$.

The corresponding function $f(x) = \Sigma c_n \phi(x - \omega_n)$ has a differential coefficient at every point not belonging to G_n. At the point $x = \omega_n$, its derivative oscillates between values $\frac{1}{2}c_n + \lambda(\omega_n)$ and $\frac{3}{2}c_n + \lambda(\omega_n)$.

2*. Let $\phi(y) = y^{\frac{1}{3}}$; then $\phi'(0) = +\infty$. The corresponding function $\Sigma c_n (x - \omega_n)^{\frac{1}{3}}$ has differential coefficients which are finite at a set of points not belonging to G. At a point ω_n of G, we have $f'(\omega_n) = +\infty$. This example does not fall under the case considered above, because $|\phi'(y)|$, for $|y| > 0$, has no upper limit.

THE CONSTRUCTION OF CONTINUOUS NON-DIFFERENTIABLE FUNCTIONS.

425. A general method of constructing functions which, although they are continuous, possess at no point a finite differential coefficient, has been developed by Dini†. It will be sufficient to discuss here a restricted class of such functions, which exhibits all the features of the somewhat more general case considered by Dini.

Let $u_1(x), u_2(x), \ldots u_n(x), \ldots$ be functions each of which is continuous in an interval (a, b), and which are such that the series $\sum\limits_1^\infty u_n(x)$ converges everywhere in (a, b), and defines a continuous function. It will be further assumed that $u_n(x)$, for each value of n, possesses maxima and minima such that the interval between each maximum and the next minimum is a number δ_n which diminishes indefinitely as n is indefinitely increased; and also that $u_n(x) = -u_n(x + \delta_n)$, so that all the maxima of $u_n(x)$ are equal to one another, and also all the minima, the maxima and minima being equal in absolute value, and opposite in sign. Let D_n denote the excess of a maximum over a minimum. It will also be assumed that $u_n(x)$ possesses finite differential coefficients of the first and second orders $u_n'(x)$, $u_n''(x)$ everywhere in (a, b); and that the upper limits of $|u_n'(x)|$, $|u_n''(x)|$ in (a, b), have finite values $\bar{u}$, $\bar{\bar{u}}$.

We have, for any two points x, $x + h$, in (a, b),

$$\frac{f(x+h) - f(x)}{h} = \frac{u_m(x+h) - u_m(x)}{h} + \sum_{n=1}^{m-1} \frac{u_n(x+h) - u_n(x)}{h} + \frac{R_m(x+h) - R_m(x)}{h},$$

where $R_m(x)$ denotes the remainder after m terms of the series $\Sigma u_n(x)$.

* This function $\Sigma c_n (x - \omega_n)^{\frac{1}{3}}$ has been studied by Brodén, see his paper "Ueber das Weierstrass-Cantor'sche Condensationverfahren," *Stockholm Öfv.*, 1896, p. 583; also *Math. Annalen*, vol. LI. See further Pompeiu, *Math. Annalen*, vol. LXIII, p. 326, where it is shewn that, if the series be denoted by t, the inverse function $x = G(t)$ is a continuous function with a limited differential coefficient which is zero at an everywhere-dense set of points, provided the series $\Sigma c_n^{\frac{1}{3}}$ be convergent. This function is accordingly everywhere-oscillating.

† *Annali di Mat.* ser. 2, vol. VIII; also *Grundlagen*, p. 206. See also Lerch, *Crelle's Journal*, vol. CIII, and Darboux, *Annales de l'école normale*, ser. 2, vol. VIII.

This equation may be written in either of the following forms:—

$$\frac{f(x+h)-f(x)}{h}=\frac{u_m(x+h)-u_m(x)}{h}+\eta_m\sum_{n=1}^{m-1}\bar{u}_n+\frac{R_m(x+h)-R_m(x)}{h}; \quad (1)$$

since $|u_n'(x)|$, $\left|\frac{u_n(x+h)-u_n(x)}{h}\right|$ have the same upper limits in (a, b); the number η_m lies between 1 and -1. Also

$$\frac{f(x+h)-f(x)}{h}=\frac{u_m(x+h)-u_m(x)}{h}+\sum_{n=1}^{m-1}u_n'(x)+\zeta_m\frac{h}{2}\sum_{n=1}^{m-1}\bar{\bar{u}}_n+\frac{R_m(x+h)-R_m(x)}{h}, \quad \ldots(2)$$

where ζ_m is a number between 1 and -1.

Let a neighbourhood $(x, x+\epsilon)$, or $(x-\epsilon, x)$ on either side of x be taken; m may then be chosen so great, that several oscillations of $u_m(x)$ are completed in the chosen neighbourhood. Let the point $x+h$ be taken at a maximum or minimum of $u_m(x)$ in $(x, x+\epsilon)$, or in $(x-\epsilon, x)$; and let it be at the first maximum or minimum of $u_m(x)$ on the right or on the left of x, of which the distance from x is $\geqq \frac{1}{2}\delta_m$. The condition

$$|u_m(x+h)-u_m(x)| \geqq \tfrac{1}{2}D_m$$

is then satisfied; also $|h| \leqq \frac{3}{2}\delta_m$. We may write

$$u_m(x+h)-u_m(x)=\tfrac{1}{2}\alpha_m\gamma_m D_m,$$

where γ_m is positive and >1, and $\alpha_m=\pm 1$, its sign depending on x and m, and possibly on the sign of h.

The equation (1) may now be written in the form

$$\frac{f(x+h)-f(x)}{h}=\frac{\alpha_m\gamma_m D_m}{2h}\left[1+\frac{2\eta_m' h}{D_m}\sum_{m=1}^{n-1}\bar{u}_n+\frac{2\alpha_m}{\gamma_m}\cdot\frac{R_m(x+h)-R_m(x)}{D_m}\right], \ldots\ldots(3)$$

where η_m' is between 1 and -1.

Next, let $x+h_1$ be the next following extreme point of $u_m(x)$ after $x+h$, so that h and h_1 have the same sign and $|h_1|>|h|$. The difference $u_m(x+h_1)-u_m(x)$, when it is not zero, has the sign opposite to that of $u_m(x+h)-u_m(x)$, and therefore

$$u_m(x+h_1)-u_m(x)=-\tfrac{1}{2}\epsilon_m'\alpha_m\gamma_m D_m,$$

where $0 \leqq \epsilon_m' < 1$.

In (1) and (2) we may write h_1 instead of h, provided the values of η_m and ζ_m are changed; we find therefore from (3),

$$\frac{f(x+h)-f(x)}{h}-\frac{f(x+h_1)-f(x)}{h_1}=\frac{\alpha_m\gamma_m D_m}{2h}\left[1+\epsilon_m'\frac{h}{h_1}+\frac{4h\eta_m''}{D_m}\sum_{n=1}^{m-1}\bar{u}_n+2\frac{\alpha_m}{\gamma_m}\frac{R_m(x+h)-R_m(x)}{D_m}-2\frac{\alpha_m}{\gamma_m}\frac{h}{h_1}\frac{R_m(x+h_1)-R_m(x)}{D_m}\right], \ldots\ldots\ldots(4)$$

where η_m'' is between 1 and -1. Also from (2), we find

$$\frac{f(x+h)-f(x)}{h} - \frac{f(x+h_1)-f(x)}{h_1}$$

$$= \frac{\alpha_m \gamma_m D_m}{2h}\left[1+\epsilon_m'\frac{h}{h_1}+\zeta_m'\frac{h(h+h_1)}{D_m}\sum_{n=1}^{m-1}\bar{u}_n + 2\alpha_m\theta_m\frac{R_m(x+h)-R_m(x)}{D_m}\right.$$

$$\left. - 2\alpha_m\theta_m'\cdot\frac{h}{h_1}\frac{R_m(x+h_1)-R_m(x)}{D_m}\right], \quad \ldots\ldots\ldots(5)$$

when ζ_m' is between 1 and -1, and θ_m, θ_m' are each between 0 and 1.

However small the neighbourhood $(x, x+\epsilon)$, or $(x-\epsilon, x)$, be chosen, m can be chosen so great that $x+h$, $x+h_1$ both lie in that one of these neighbourhoods which is under consideration. As ϵ is indefinitely diminished, m must be indefinitely increased; thus, on passing to the limit, the results (3), (4), (5) may be applied to obtain conditions for the existence or the non-existence of differential coefficients of the function $f(x)$.

(*a*) Let it be assumed that, as m is increased, the distance between successive extremes of $u_m(x)$ becomes continually smaller compared with D_m, and so that $\frac{\delta_m}{D_m}$, and therefore also $\frac{h}{D_m}$, has the limit zero when m is indefinitely increased.

In this case the limit of $\frac{f(x+h)-f(x)}{h}$ is certainly infinite, unless the expression in the bracket in (3) has the limit zero.

If then $R_m(x+h)-R_m(x)$ has, for values of m greater than an arbitrarily chosen integer m', the same sign as α_m, *i.e.* as $u_m(x+h)-u_m(x)$, and if further $\frac{2h}{D_m}\sum_{m=1}^{n-1}\bar{u}_m$ remains in absolute value less than unity, by more than some fixed difference, $\frac{f(x+h)-f(x)}{h}$ cannot have a finite limit. The last condition may be replaced by the condition, that $\frac{3\delta_m}{D_m}\sum_{m=1}^{n-1}\bar{u}_n$ must remain numerically less than unity, by more than some fixed difference.

If these conditions are satisfied, one at least of the derivatives $D^+f(x)$, $D_+f(x)$ is infinite, and also one at least of $D^-f(x)$, $D_-f(x)$ is infinite. Moreover, in the case of functions of the type here considered,

$$u_m(x+h)-u_m(x)$$

has the same sign for positive as for negative values of h, hence it is impossible that $f(x)$ can have an infinite differential coefficient with a fixed sign; and therefore $f(x)$ has at no point a differential coefficient, either finite or infinite.

(*b*) If it be not known that the condition in (*a*) relating to the sign of $R_m(x+h) - R_m(x)$ is satisfied, then $f(x)$ has no differential coefficients, provided $|R_m(x+h) - R_m(x)|$ has a finite upper limit $2R_m'$ for all values of x, and also

$$\frac{3\delta_m}{D_m}\sum_{m=1}^{n-1}\bar{u}_n + \frac{4R_m'}{D_m}$$

remains less than unity, by more than some fixed difference. The condition as to the limit of $\frac{\delta_m}{D_m}$ is the same as in (*a*).

(*c*) Let us next suppose that $\frac{\delta_m}{D_m}$ has not the limit zero, but remains less than some finite number, for all values of m. In this case $\frac{D_m}{h}$ has a finite limit; and we see from (4) or (5), that

$$\frac{f(x+h)-f(x)}{h} - \frac{f(x+h_1)-f(x)}{h_1}$$

has not the limit zero, provided the expression in the bracket, in either case, does not converge to zero.

If then $R_m(x+h) - R_m(x)$ has the same sign as $u_m(x+h) - u_m(x)$, and $R_m(x+h_1) - R_m(x)$ has the opposite sign, and if further one of the two expressions

$$\frac{6\delta_m}{D_m}\sum_{n=1}^{m-1}\bar{u}_n,$$

$$\frac{6\delta_m^2}{D_m}\sum_{n=1}^{m-1}\bar{\bar{u}}_n$$

remains less than unity, by more than some fixed difference, then $f(x)$ can nowhere have a finite differential coefficient, although it may have an infinite one, at some points.

(*d*) The same condition relating to $\frac{\delta_m}{D_m}$ holding, as in (*c*), it is sufficient to ensure that $f(x)$ has nowhere a finite differential coefficient, that

$$|R_m(x+h) - R_m(x)|, \quad |R_m(x+h_1) - R_m(x)|$$

should never exceed a fixed number $2R_m'$, and that one of the two expressions

$$\frac{6\delta_m}{D_m}\sum_{n=1}^{m-1}\bar{u}_n + \frac{32}{5}\frac{R_m'}{D_m}, \quad \frac{6\delta_m^2}{D_m}\sum_{n=1}^{m-1}\bar{\bar{u}}_n + \frac{32}{5}\frac{R_m'}{D_m}$$

should remain less than unity, by more than some fixed difference. These expressions are obtained from equations (4) and (5), by taking account of the fact that h/h_1 cannot exceed 3/5.

426. Let $v_n(x)$ be a continuous function which has maxima and minima equal to 1 and -1 respectively, at distances d_n from one another; where d_n is, for every value of n, less than some fixed number. Also suppose that $v_n(x) = -v_n(x+d_n)$, and that $v_n'(x)$, $v_n''(x)$ are, for all values of x, in absolute value not greater than some fixed positive number A. We then take the function $u_n(x)$ of § 425 to be $a_n v_n(b_n x)$, where the a_n's are constants such that $\Sigma |a_n|$ is a convergent series, and the b_n's are constants such that $a_n b_n$ does not converge to zero as n is indefinitely increased. The convergency condition of $\Sigma |a_n|$ ensures that the series $\Sigma a_n v_n(b_n x)$ converges uniformly, and thus that $f(x)$ is a continuous function.

We have in this case

$$D_m = 2|a_m|, \quad \delta_m = \frac{d_m}{b_m};$$

also $\quad |R_m(x+h) - R_m(x)| \leqq 2R_m', \quad |R_m(x+h_1) - R_m(x)| \leqq 2R_m',$

where R_m' is the remainder of the series $\Sigma |a_n|$ after m terms.

We see from (*b*) in § 425, that if $a_m b_m$ becomes indefinitely great with m, then $f(x) = \Sigma a_n v_n(b_n x)$ has nowhere a differential coefficient, in case the condition

$$\frac{3}{2}\frac{Ad_m}{b_m|a_m|}\sum_1^{m-1} b_n|a_n| + \frac{2R_m'}{|a_m|} < 1,$$

by more than some fixed difference, be satisfied.

In case $a_m b_m$ does not necessarily become indefinitely great with m, but has not the limit zero, we see from (*d*) of § 425, that $f(x)$ is not differentiable, provided one of the conditions

$$\frac{3Ad_m}{b_m|a_m|}\sum_1^{m-1} b_n|a_n| + \frac{16}{5}\frac{R_m'}{|a_m|} < 1, \quad \frac{3Ad_m^2}{b_m^2|a_m|}\sum_1^{m-1} b_n^2|a_n| + \frac{16}{5}\frac{R_m'}{|a_m|} < 1$$

is satisfied in the limit.

Let us now consider the special case in which d_n is constant, and $= d$, and let all the functions $v_n(x)$ have their maxima and minima at the points $0, \pm d, \pm 2d \dots$ Further let us suppose that b_{n+s}/b_n is, for every value of s, an odd integer. Then, if $b_m(x+h)$ corresponds to a maximum or minimum of $v_m(x)$, $b_{m+s}(x+h)$ also corresponds to a maximum or minimum of $v_{m+s}(x)$. If we suppose all the a_n's to be positive, then the difference

$$R_m(x+h) - R_m(x) = \sum_{m+1}^{\infty} a_n\{v_n[b_n(x+h)] - v_n(b_n x)\}$$

is either zero, or else it has the same sign as $u_m(x+h) - u_m(x)$; on the other hand the difference $R_m(x+h_1) - R_m(x)$ has the opposite sign. We can consequently apply the criteria in (*a*) and (*c*) of § 425. The sufficient conditions that $\Sigma a_n v_n(b_n x)$ is not differentiable become in this case,

$$\frac{3}{2}\frac{Ad}{a_m b_m}\sum_1^{m-1} a_n b_n < 1, \quad \lim a_m b_m = \infty;$$

and when $a_m b_m$ has not the limit zero, then one of the conditions

$$\frac{3Ad}{a_m b_m}\sum_{n=1}^{m-1} a_n b_n < 1, \quad \frac{3Ad^2}{a_m b_m^2}\sum_{n=1}^{m-1} a_n b_n^2 < 1,$$

by less than a fixed number, must be satisfied.

In particular, let $a_n = a^n$, $a < 1$, $b_n = b^n$, where b is a positive odd integer.

If $ab > 1$, the condition of non-differentiability is satisfied if

$$\frac{3}{2}\frac{Ad}{a^n b^n} \cdot \frac{a^n b^n - 1}{ab - 1} < 1,$$

for every value of n; and this reduces to $ab > 1 + \frac{3}{2}Ad$.

If $ab \geqq 1$, the condition of non-differentiability is $ab^2 > 1 + 3Ad^2$.

If we let $v_n(x) = \cos x$, we obtain Weierstrass' theorem* that, if b is an odd integer, the continuous function

$$\sum_1^\infty a^n \cos b^n x \quad (a < 1)$$

is not differentiable for any value of x, provided either $ab > 1 + \frac{3}{2}\pi$, or else provided $ab^2 > 1 + 3\pi^2$, and also $ab \geqq 1$. In the first case $ab > 1 + \frac{3}{2}\pi$, there can be no infinite differential coefficient with a fixed sign, although at some points there may exist infinite derivatives on the two sides with opposite signs. This was the first example of a continuous function nowhere differentiable which was exhibited.

EXAMPLES.

1. The function represented by $\sum_1^\infty a^n \sin b^n x$, $(a < 1)$, where b is an integer of the form $4p+1$, is continuous, but nowhere has a finite differential coefficient, provided one of the two conditions $ab > 1 + \frac{3}{2}\pi$, or $ab^2 > 1 + 3\pi^2$, is satisfied, and $ab \geqq 1$.

2. The functions† represented by

$$\Sigma \frac{1}{31^n} \cos(31^n x), \quad \Sigma \frac{1}{33^n} \cos(33^n x), \quad \Sigma \frac{1}{33^n} \sin(33^n x),$$

do not possess finite differential coefficients.

3. The continuous functions represented by

$$\sum_1^\infty \frac{a^n}{1.3.5\ldots(2n-1)} \cos\{1.3.5\ldots(2n-1)x\},$$

$$\sum_1^\infty \frac{a^n}{1.5.9\ldots(4n+1)} \sin\{1.5.9\ldots(4n+1)x\},$$

where $a > 1 + \frac{3}{2}\pi$, do not possess finite differential coefficients.

* *Crelle's Journal*, vol. LXXIX.

† See Wiener, *Crelle's Journal*, vol. XC.

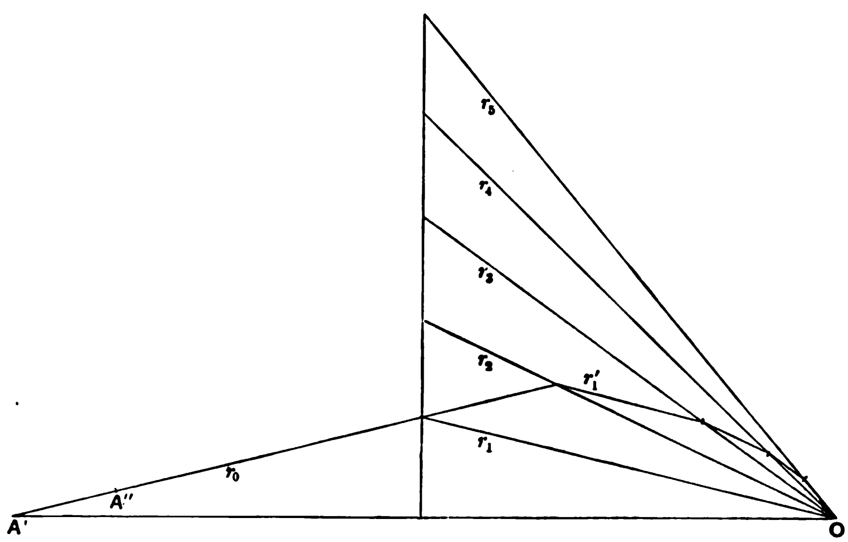
THE CONSTRUCTION OF A DIFFERENTIABLE EVERYWHERE-OSCILLATING FUNCTION.

427. The first attempt to construct a function with maxima and minima in every interval, which should have at every point a finite differential coefficient, was made by Hankel*. The function which he constructed is however not an everywhere-oscillating function. By Du Bois Reymond† the view was expressed that no such function can exist, but Dini‡ regarded the existence of such functions as highly probable. The first actual construction of such a function is due to Köpcke, who having first§ constructed an everywhere-oscillating function with derivatives on the right and on the left at every point, in a subsequent memoir|| obtained a function having the required properties. Köpcke's construction has been simplified by Pereno¶, and the account here given is based upon the work of the latter.

On a straight line AB measure off segments AA', $B'B$, each equal to $\frac{1}{20}AB$. Let O be the middle point of AB, and draw through O straight lines $r_1, r_2, r_3, \dots r_{2^n+1}$, making angles with OA of which the tangents are

$$1/2^n,\quad 2/2^n,\quad 3/2^n, \dots (2^n+1)/2^n$$

respectively. Through A' draw a straight line r_0 making with $A'O$ an angle of tangent $1/2^n$. Through the intersection (r_0, r_2) of r_0 and r_2, draw a straight line r_1' parallel to r_1: through (r_1', r_3) draw a straight line r_2' parallel to r_2, and so on. The straight lines $r_0, r_1', r_2', \dots r_{2^n-1}, r_{2^n+1}$ form an unclosed polygon



* *Math. Annalen*, vol. XX.

† *Crelle's Journal*, vol. LXXIX.

‡ *Grundlagen*, p. 383.

§ *Math. Annalen*, vol. XXIX.

|| *Math. Annalen*, vols. XXXIV. and XXXV.

¶ *Giorn. di Mat.*, vol. XXXV, 1897.

above $A'O$. On OB' describe a precisely similar polygon on the other side of AB. The figure is drawn for the case $n=2$, and shews the half of the figure belonging to $A'O$. The two polygons form a single polygon joining $A'B'$, and crossing it at O. On r_0 take $A'A''=AA'$, and describe an arc of a circle touching AB at A, and r_0 at A''. At each vertex of the polygon which has been constructed, mark off on the sides adjacent to that vertex lengths equal to $\frac{1}{30}$ of the shorter side, and construct an arc of a circle touching the two sides at the extremities of these segments so marked off. We have now a figure joining A and B, and composed of arcs of circles and of straight lines. This figure, by means of its ordinates perpendicular to AB, defines a continuous differentiable function, with a continuous differential coefficient which is zero at A and B, and is $-(2^n+1)/2^n$ at O. This function may be denoted by $(A/B)_n$.

Let x, y be a system of coordinate axes in a plane, and draw a quadrant of a circle passing through the points (0, 0) and (1, 0), in the positive quadrant. Let $F_0(x)$ be the function represented by this quadrant, for the interval (0, 1) of x. The function $F_0(x)$ has a maximum at $x=\frac{1}{2}$; also $F_0'(0)=1$, $F_0'(1)=-1$. If a_0 denote the value of $F_0'(x)$ at $x=\frac{1}{4}$, describe the curve of which the ordinates are $a_0(0\,|\,\frac{1}{2})$, from $x=0$ to $x=\frac{1}{2}$, and $-a_0(\frac{1}{2}\,|\,1)$, from $x=\frac{1}{2}$ to $x=1$. This curve represents a continuous function $f_1(x)$; and we have

$$f_1'(0)=f_1'(\tfrac{1}{2})=f_1'(1)=0,$$

and

$$f_1'(\tfrac{1}{4})=-\tfrac{3}{2}a_0,\quad f_1'(\tfrac{3}{4})=\tfrac{3}{2}a_0.$$

The function

$$F_1(x)=F_0(x)+f_1(x)$$

is such that

$$F_1'(\tfrac{1}{4})=-\tfrac{1}{2}a_0,\quad F_1'(\tfrac{3}{4})=\tfrac{1}{2}a_0,\quad F_1'(0)=1,\quad F_1'(1)=-1,\quad F_1'(\tfrac{1}{2})=0.$$

Thus $F_1(x)$ has a maximum in the interval $(0, \frac{1}{4})$, a minimum in $(\frac{1}{4}, \frac{1}{2})$, a maximum at $x=\frac{1}{2}$, a minimum in $(\frac{1}{2}, \frac{3}{4})$, and a maximum in $(\frac{3}{4}, 1)$.

Let the interval (0, 1) be divided into sub-intervals, by means of the points at which $F_1'(x)=0$; then in each of these sub-intervals $F_1(x)$ is monotone. Then divide each of these sub-intervals into 2, 4, 8, ... equal parts, until the fluctuation of $F_1'(x)$ in each of these parts is $\leq\frac{1}{2}$: this is always possible, since $F_1'(x)$ is a continuous function. Let $c_1^{(1)}, c_1^{(2)}, c_1^{(3)}, \ldots$ denote all the points in which (0, 1) has been divided in this manner. In any one part $(c_1^{(s-1)}, c_1^{(s)})$, $F_1(x)$ is monotone, and its differential coefficient has a fluctuation $\leq\frac{1}{2}$. Let $a_1^{(1)}, a_1^{(2)}, \ldots$ denote the values of $F_1'(x)$ at the middle points of the intervals $(0, c_1^{(1)}), (c_1^{(1)}, c_1^{(2)}), \ldots$. Describe the curves

$$a_1^{(1)}(0\,|\,c_1^{(1)})_2,\quad a_1^{(2)}(c_1^{(1)}\,|\,c_1^{(2)})_2,\quad a_1^{(3)}(c_1^{(2)}\,|\,c_1^{(3)})_2\ldots;$$

these form together a continuous curve which represents a function $f_2(x)$. Let

$$F_2(x)=F_0(x)+f_1(x)+f_2(x);$$

then $F_2(x)$ has, in every interval $(c_1^{(s-1)}, c_1^{(s)})$, a new maximum and a new minimum. The length of each interval is $< 1/2^2$.

Proceeding in this manner, let us suppose that the function $F_n(x)$ has been formed. Take the points at which $F_n'(x)$ vanishes, and, in case $F_n(x)$ has lines of invariability, the limiting points of those lines; these points divide (0, 1) into sub-intervals in each of which $F(x)$ is monotone. Then divide each of these sub-intervals into 2, 4, 8, ... parts, until the fluctuation of $F_n'(x)$ in each part is $\leq 1/2^n$; let $c_n^{(1)}, c_n^{(2)}, c_n^{(3)} \ldots$ be all the points of division of (0, 1) thus formed. In any interval $(c_n^{(s-1)}, c_n^{(s)})$, the function $F_n(x)$ is monotone, and the fluctuation of $F_n'(x)$ is $\leq 1/2^n$. Let

$$a_n^{(1)}, a_n^{(2)}, \ldots a_n^{(s)}, \ldots$$

be the values of $F_n'(x)$ at the middle points of the intervals; and in the case of a line of invariability, take as the corresponding value of the a_n, $1/2^n$ or $-1/2^n$, according as the line of invariability is in the interval $(0, \frac{1}{2})$, or in $(\frac{1}{2}, 1)$. Let the curves $a_n^{(s)}(c_n^{(s-1)} \mid c_n^{(s)})_{n+1}$ be described, and let the function represented by the totality of these curves be denoted by $f_{n+1}(x)$. Then the function

$$F_{n+1}(x) = F_n(x) + f_{n+1}(x)$$

has a new maximum, and a new minimum, in every interval $(c_n^{(s-1)}, c_n^{(s)})$, and the length of each of these intervals is less than $1/2^{n+1}$.

If this law of generation of the functions $f_n(x)$ be employed indefinitely, we have a series

$$F_0(x) + f_1(x) + f_2(x) + \ldots + f_n(x) + \ldots;$$

and it will be shewn that this series represents a continuous function which is everywhere differentiable, and which has an everywhere-dense set of maxima and minima.

428. Let

$$F_0'(x) + f_1'(x) + f_2'(x) + \ldots + f_n'(x) = S_n(x);$$

it will then be shewn that, for every value of n and x, $S_n(x)$ is numerically less than $\prod_{n=1}^{\infty}\left(1+\frac{1}{2^n}\right)$, which may be denoted by P. Let us assume that $|S_n(x)|$ is, for every value of x, less than $\prod_{1}^{n}\left(1+\frac{1}{2^n}\right)$, which may be denoted by P_n: it will then be shewn that $|S_{n+1}(x)| < P_{n+1}$.

Let the point x be in the interval $(c_n^{(s-1)}, c_n^{(s)})$, where $x < c_n^{(s)}$, the number s depending upon the value of x; we have then, in accordance with the construction of the functions,

$$S_{n+1}(x) = S_n(x) + a_n \frac{a_n^{(s)}}{2^{n+1}},$$

where $$1 \geqq a_n \geqq -(2^{n+1}+1).$$

In the interval $(c_n^{(s-1)}, c_n^{(s)})$, $S_n(x)$ has a fixed sign, the same as that of $a_n^{(s)}$, but this is not the case for $S_{n+1}(x)$. If a_n is positive, we have

$$|S_{n+1}(x)| < P_n\left(1+\frac{1}{2^n}\right) < P_{n+1}.$$

If a_n is negative, we have

$$|S_{n+1}(x)| < |S_n(x)| < P_n < P_{n+1};$$

it has thus been shewn that if $|S_n(x)| < P_n$, then also $|S_{n+1}(x)| < P_{n+1}$. Now $|F_1'(x)|$ is, everywhere in (0, 1), less than $(1+\frac{1}{2})$, and therefore the theorem $|S_n(x)| < P_n$ follows by induction. *A fortiori* $|S_n(x)|$ is, for every value of n and x, $< P$.

The numerically greatest value of $f_{n+1}(x)$ in the interval $(c_n^{(s-1)}, c_n^{(s)})$ is at some point on the left of the middle point of the interval, and that value is consequently $< \frac{1}{2}\cdot\frac{1}{2^{n+1}}\cdot\frac{a_n^{(s)}}{2^{n+1}}$, since the length of the interval is less than $1/2^{n+1}$. Also, as has been shewn above, $a_n^{(s)} < P$, and therefore

$$|f_{n+1}(x)| < \frac{P}{2^{2n+3}};$$

and hence, since the terms of the series $f_1(x)+f_2(x)+\dots$ are numerically less than the corresponding terms of the absolutely convergent series

$$\frac{P}{2^5}+\frac{P}{2^7}+\dots+\frac{P}{2^{2n+3}}+\dots,$$

it follows that the series $f_1(x)+f_2(x)+\dots$ is uniformly convergent in the interval (0, 1). It follows that the function $F(x)$ defined as the sum-function of the series $F_0(x)+f_1(x)+f_2(x)+\dots$ is a continuous function.

In order to prove that the function $F(x)$ is everywhere differentiable, we shall shew that it satisfies the conditions stated in the theorem of § 398.

We have first of all to shew that the series $f_1'(x)+f_2'(x)+\dots$ is convergent for all values of x in (0, 1). In case, for any value of x, all the numbers $S_n(x), S_{n+1}(x), \dots$, from and after some value of n, have all the same sign, say the positive sign, we have

$$S_{m+1}(x) \leqq S_m(x) + \frac{a_m^{(s_1)}}{2^{m+1}},$$

where m is the value of n in question. Also

$$S_{m+2}(x) \leqq S_{m+1}(x) + \frac{a_{m+1}^{(s_2)}}{2^{m+2}},$$

with similar inequalities involving higher indices. From these inequalities, we find

$$S_{m+p}(x) - S_m(x) \leqq \frac{a_m^{(s_1)}}{2^{m+1}} + \frac{a_{m+1}^{(s_2)}}{2^{m+2}} + \dots + \frac{a_{m+p-1}^{(s_p)}}{2^{m+p}} \leqq \frac{P}{2^m};$$

and since m may be taken so great that $P/2^m$ is arbitrarily small, we see that m may be so chosen that $S_{m+p} - S_m(x)$ is arbitrarily small, whatever positive integral value p may have. It has thus been shewn that, in the case considered, the series is convergent.

It may happen that $S_n(x)$ is zero, owing to x being at a point of division $a_n^{(s)}$; in this case all the functions $f_n'(x)$ with higher indices vanish, and therefore all the functions $S_n(x)$ vanish, from and after the particular value of n. It may happen that $S_n(x)$ vanishes, owing to x being a point of invariability of $F_n(x)$; in this case $S_{n+1}(x)$ may vanish if x is an extreme of $f_{n+1}(x)$, and then x is a point of division $a_{n+1}^{(s)}$, and all the functions $S_m(x)$ for indices $m > n$ vanish. Thus if, for any value of x, $S_n(x)$, $S_{n+1}(x)$ both vanish, then $S_m(x)$ vanishes for all values of $m \geqq n$. If $S_n(x)$ vanishes, but not $S_{n-1}(x)$ or $S_{n+1}(x)$, x is a point of invariability of $F_n(x)$, and

$$S_{n+1}(x) \leqq \frac{1}{2^n}\left(1 + \frac{1}{2^{n+1}}\right) < \frac{2P}{2^{n+1}},$$

and the same reasoning is applicable as before. Let us next suppose that the functions $S_n(x)$ are never all of the same sign, from and after any value n, and that for some values of n they vanish; let $n_1, n_2, \ldots$ be the values of n for which $S_n(x)$ has a change of sign, for example, let $S_{n_1}(x)$ be negative or zero, and $S_{n_1+1}(x)$ be positive, and $S_{n_2}(x)$ positive or zero, and $S_{n_2+1}(x)$ negative, and so on. If $S_{n_1}(x)$ is negative, we have

$$S_{n_1+1}(x) = S_{n_1}(x) + \frac{a_{n_1}^{(s_1)}}{2^{n_1+1}} \alpha_{n_1},$$

where
$$1 \geqq \alpha_{n_1} \geqq -(2^{n_1+1} + 1),$$
and since α_{n_1} is negative, we have

$$S_{n_1+1}(x) < \frac{1}{2^{n_1}} + \frac{P}{2^{n_1+1}} < \frac{P}{2^{n_1}},$$

account being taken of the fact that the fluctuation of $F_{n_1}'(x)$ in the interval in which x lies is $\leqq \frac{1}{2^{n_1}}$. If $S_{n_1}(x)$ is zero, so that x is a point of invariability of $F_{n_1}(x)$, we have

$$S_{n_1+1}(x) \leqq \frac{1}{2^{n_1+1}}\left(1 + \frac{1}{2^{n_1+1}}\right) < \frac{P}{2^{n_1}}.$$

In any case we find that

$$S_{n_1+p}(x) < S_{n_1+1}(x) + \frac{P}{2^{n_1+1}} < \frac{P}{2^{n_1}} + \frac{P}{2^{n_1+1}},$$

where
$$p = 1, 2, \ldots n_2 - n_1.$$
Similarly, we find that

$$S_{n_2+1}(x) < \frac{P}{2^{n_2}}, \text{ if } S_{n_2}(x) = 0;$$

and if $S_{n_2}(x) > 0$, we have

$$|S_{n_2+p}(x)| < |S_{n_2+1}(x)| + \frac{P}{2^{n_1+1}} < \frac{P}{2^{n_1}} + \frac{P}{2^{n_1+1}},$$

for $$p = 1, 2, 3, \ldots n_3 - n_2.$$

It is seen from these results that $|S_n(x)|$ becomes arbitrarily small for all sufficiently great values of n, and thus $\lim_{n=\infty} S_n(x) = 0$. It has now been shewn that in every case the series

$$F_0'(x) + f_1'(x) + f_2'(x) + \ldots$$

converges for each value of x in the interval $(0, 1)$.

429. It must next be proved that, if ϵ be an arbitrarily chosen positive number, then, for a given x, a number $\delta > 0$ can be found, such that, for each value of h numerically less than δ, and for which $x + h$ is in the interval $(0, 1)$, there exists an integer m, variable with h, and not less than a prescribed integer m', such that the three numbers

$$\frac{F_m(x+h) - F_m(x)}{h} - S_n(x), \quad \frac{R_m(x+h)}{h}, \quad \frac{R_m(x)}{h}$$

are all numerically less than ϵ; $R_m(x)$ denoting the remainder of the series which represents $F(x)$, that is, $F(x) - F_{m-1}(x)$.

The case may be left out of account in which x coincides with one of the points of division of $(0, 1)$; for the function $F(x)$ is then represented by a finite series, and is differentiable, since $f'_{n+p}(c_n^{(s)}) = 0$, for $p > 1$.

Let ϵ, m' be fixed, and let us consider a point x in $(0, 1)$; then a number $n \geqq m'$ can be so determined that

$$\frac{P}{2^{n-2}} < \frac{1}{3}\epsilon, \text{ and } |S_{n+p}(x) - S_{n+q}(x)| < \frac{1}{3}\epsilon,$$

where p, q are any positive integers. For any value of h, such that $x + h$ falls within the interval $(c_n^{(s-1)}, c_n^{(s)})$, the number m can be determined. Let h be positive, and determine n_1 so that $x < c_{n_1+1}^{(s')} \leqq x + h \leqq c_{n_1}^{(s'')} \leqq c_n^{(s)}$; then it can be shewn that $n_1 + 2$ is a suitable value for m. We have

$$f_{n_1+1+p}(c_{n_1+1}^{(s)}) = 0;$$

and $$|f_{n_1+1+p}(c_{n_1+1}^{(s')} \pm k)| < \frac{P}{2^{n_1+1+p}}k, \text{ for } p \geqq 1.$$

The point $c_{n_1+1}^{(s')}$ is in general between x and $x + h$, and therefore it determines two segments, k_1, k_2, where

$$x = c_{n_1+1}^{(s')} - k_1, \quad x + h = c_{n_1+1}^{(s')} + k_2.$$

We have therefore

$$|f_{n_1+1+1}(x)| < \frac{P}{2^{n_1+1+1}}k_1, \quad |f_{n_1+1+2}(x)| < \frac{P}{2^{n_1+1+2}}k_1.$$

and so on; and from these inequalities we find that

$$|R_{n_1+2}(x)| < Pk_1\left\{\frac{1}{2^{n_1+2}} + \frac{1}{2^{n_1+3}} + \dots\right\} < \frac{Pk_1}{2^{n_1+1}},$$

and similarly that

$$|R_{n_1+2}(x+h)| < \frac{Pk_2}{2^{n_1+1}}.$$

Since k_1, k_2 are less than h, we have

$$\left|\frac{R_{n_1+2}(x)}{h}\right| < \frac{P}{2^{n_1+1}} < \epsilon, \text{ and } \left|\frac{R_{n_1+2}(x+h)}{h}\right| < \frac{P}{2^{n_1+1}} < \epsilon.$$

It has thus been shewn that $m = n_1 + 2$ is a value of m which satisfies the required condition. The case in which h is negative can be treated in the same manner.

We have now to prove that

$$\left|\frac{F_{n_1+1}(x+h) - F_{n_1+1}(x)}{h} - S_{n+2}(x)\right| < \epsilon.$$

We see that

$$\frac{F_{n_1+1}(x+h) - F_{n_1+1}(x)}{h} - S_{n+2}(x)$$
$$= \left\{\frac{F_{n_1}(x+h) - F_{n_1}(x)}{h} - S_{n_1+1}(x)\right\} + \left\{\frac{f_{n_1+1}(x+h) - f_{n_1}(x)}{h} - f'_{n_1+1}(x)\right\};$$

and if x, $x+h$ are points in $(c_{n_1}^{(s-1)}, c_{n_1}^{(s)})$, the absolute value of the first term on the right-hand side is not greater than $1/2^{n_1}$.

We consider therefore

$$\frac{f_{n_1+1}(x+h) - f_{n_1+1}(x)}{h} - f'_{n_1+1}(x).$$

From the construction for $f_{n_1+1}(x)$, we have

$$\frac{f_{n_1+1}(x+h) - f_{n_1}(x)}{h} \leq \frac{a_{n_1}^{(s)}}{2^{n_1+1}},$$

since x, $x+h$ are in the interval $(c_{n_1}^{(s-1)}, c_{n_1}^{(s)})$. Let us take the case in which $F_{n_1}(x)$ increases from $c_{n_1}^{(s-1)}$ to $c_{n_1}^{(s)}$; then, for any point x in the interval between these two points, we have

$$\frac{f_{n_1+1}(x+h) - f_{n_1+1}(x)}{h} \leq \frac{a_{n_1}^{(s)}}{2^{n_1-1}}.$$

We shall find also a lower limit for this incrementary ratio. The point x is such that the ordinate of $a_{n_1}^{(s)}(c_{n_1}^{(s-1)}, c_{n_1}^{(s)})_{n_1-1}$ is below the x-axis, and if, for that point, the differential coefficient is negative, we have

$$\frac{f_{n_1-1}(x+h) - f_{n_1-1}(x)}{h} \geq f'_{n_1+1}(x).$$

Let the sides of the rectilinear polygon which was employed in the construction of $a_{n_1}^{(s)}(c_{n_1}^{(s-1)} \mid c_{n_1}^{(s)})_{n_1+1}$ be denoted by

$$r_0', r_1', \dots r'_{2^{n_1+1}-1}, r'_{2^{n_1+1}+1}, s'_{2^{n_1+1}-1} \dots s_2', s_1', s_0',$$

where r_m' is equal and parallel to s_m'. On r_2', produced beyond (r_1', r_2'), take a segment equal to r_2'; then this segment is equal and parallel to s_2', and the line joining the end of this segment with (s_2', s_3') is parallel to r_3, and will cut r_1' in a point p_1. But s_3' is parallel to r_3, and passes through (s_2', s_3'); therefore this segment is the prolongation of s_3', and is consequently inclined to the x-axis at an angle whose tangent is $-3\frac{a_{n_1}^{(s)}}{2^{n_1+1}}$. Hence, for a point between $c_{n_1}^{(s)}$ and p_1, for which the ordinate is positive, we have

$$\frac{f_{n_1+1}(x+h) - f_{n_1+1}(x)}{h} > -3\frac{a_{n_1}^{(s)}}{2^{n_1+1}}.$$

But the greatest value of $f'_{n_1+1}(x)$, in this case, is $\frac{a_{n_1}^{(s)}}{2^{n_1+1}}$; and therefore

$$\frac{f_{n_1+1}(x+h) - f_{n_1+1}(x)}{h} > f'_{n_1+1}(x) - 4\frac{a_{n_1}^{(s)}}{2^{n_1+1}}.$$

If a point p_2 on r_3' be determined, by making a similar construction with r_3' instead of r_2', then, for every point on the arc p_1, p_2, except p_2,

$$\frac{f_{n_1+1}(x+h) - f_{n_1+1}(x)}{h} > -4\frac{a_{n_1}^{(s)}}{2^{n_1+1}}.$$

But the maximum value of the differential coefficient is, in this case, $-\frac{a_{n_1}^{(s)}}{2^{n_1+1}}$; therefore also in this case,

$$\frac{f_{n_1+1}(x+h) - f_{n_1+1}(x)}{h} > f'_{n_1+1}(x) - 4\frac{a_{n_1}^{(s)}}{2^{n_1+1}}.$$

This condition holds for every point on the curve which has a positive ordinate. It holds also for points with a negative ordinate; because for such points with a negative differential coefficient the relation

$$\frac{f_{n_1+1}(x+h) - f_{n_1+1}(x)}{h} \geq f'_{n_1+1}(x)$$

holds; and for points where the differential coefficient is positive, the expression on the left-hand is positive, and that on the right-hand is negative.

It has now been established that

$$f'_{n_1+1}(x) - 4\frac{a_{n_1}^{(s)}}{2^{n_1+1}} < \frac{f_{n_1+1}(x+h) - f_{n_1+1}(x)}{h} \leq \frac{a_{n_1}^{(s)}}{2^{n_1+1}},$$

and it has already been proved that

$$\frac{F_{n_1}(x+h)-F_{n_1}(x)}{h}=F'_{n_1}(x)+\frac{\theta}{2^{n_1}},\ \text{ where }\ 1\geqq\theta\geqq-1.$$

We now see that

$$F'_{n_1+1}(x)+\frac{\theta}{2^{n_1}}-4\frac{a^{(s)}_{n_1}}{2^{n_1+1}}<\frac{F_{n_1+1}(x+h)-F_{n_1}(x)}{h}\leqq F'_{n_1}(x)+\frac{\theta}{2^{n_1}}+\frac{a^{(s)}_{n_1}}{2^{n_1+1}},$$

and hence $$\left|\frac{F_{n_1+1}(x+h)-F_{n_1}(x)}{h}-S_{n_1+2}(x)\right|<\epsilon,$$

since $a^{(s)}_{n_1}<P$, and $P/2^{n_1-2}<\frac{1}{3}\epsilon$, and $|\theta/2^{n_1}|<P/2^{n_1}<\frac{2}{3}\epsilon$.

It has now been established that the function $F(x)$ has at every point a finite differential coefficient which is the sum of the convergent series

$$F_0'(x)+f_1'(x)+f_2'(x)+\dots.$$

Lastly, it must be proved that $F(x)$ has an everywhere-dense set of maxima and minima.

It has been shewn that, in every interval $(c^{(s-1)}_{n-1}, c^{(s)}_{n-1})$, the function $F_n(x)$ has a new maximum and a new minimum, and that the length of the interval is less than $1/2^n$. If x_0 is a maximum of $F_n(x)$, we have

$$F_n(x_0)=F_{n+1}(x_0),\ \text{ and }\ F_n'(x_0)=F'_{n+1}(x_0)=0.$$

Moreover $f_{n+1}(x)$ is negative in the neighbourhood of the point x_0, and therefore $F_{n+1}(x_0+h)-F_{n+1}(x_0)$ is negative or zero, provided $|h|$ is less than some number k. It thus appears that $F_{n+1}(x)$ has also a maximum at x_0. If x_0 is a point of invariability of $F_n(x)$, it is no longer one for $F_{n+1}(x)$, and cannot be a point of invariability of all the functions with higher indices. If x_0 is a limiting point of a line of invariability, $F_{n+1}(x)$ will have a maximum or a minimum, or else a point of inflexion at x_0. In every case $F_{n+1}(x)$ will have a maximum and a minimum in every line of invariability of $F_n(x)$. For any given interval, as small as we please, n can be determined so great that the interval contains one of the intervals $(c^{(s-1)}_{n-1}, c^{(s)}_{n-1})$ in its interior, and all the functions $F_n(x)$, $F_{n+1}(x)$, ... have maxima in this interval; and it follows that $F(x)$ also has maxima therein.

It may be remarked that $F'(x)$, although definite at every point, has discontinuities of the second kind at an everywhere-dense set of points. At every point of continuity, this differential coefficient must vanish (see § 223). The function $F'(x)$ is not integrable in accordance with Riemann's definition.

CHAPTER VII.

TRIGONOMETRICAL SERIES.

430. THE theory of the representation of functions of a real variable by means of series of cosines and sines of multiples of the variable is of the highest importance, not only on account of the fact that such mode of representation is at present an indispensable tool in the various branches of Mathematical Physics, but also because this theory has exercised the most far-reaching influence upon the development of modern Mathematical Analysis. Historically, the questions which have arisen in connection with this theory have influenced the development of the theory of functions of a real variable to an extent which is comparable with the degree in which the theory of functions in general has been affected by the theory of power series. The theory of sets of points, which led later to the abstract theory of aggregates, arose directly from questions connected with trigonometrical series. The precise formulation by Riemann of the conception of the definite integral, the gradual development of the modern notion of a function as existent independently of any special mode of representation by an analytical expression, are further examples of the results of the study of the properties of these series upon Mathematical Analysis.

It is a significant fact that the theory of this mode of representation of a function had its origin in the attempt to investigate the form of a stretched string in a state of vibration. The problem of the expansion of the reciprocal of the distance between two planets in a series of cosines of multiples of the angle between their radii vectores led to an independent development* of the theory of trigonometrical series. The discussions which arose in connection with the first of these problems were, however, of much greater importance in the history of the development of the theory of functions; they form the first stage in the development of what is known as the theory of Fourier's series, in intimate connection with which the modern theory of functions of real variables had its origin.

* The importance of this fact has been emphasized by H. Burkhardt in his work "Entwickelungen nach oscillirenden Functionen," published as a *Jahresbericht der deutschen Mathematiker-Vereinigung*, vol. x, 1901 and later.

THE PROBLEM OF VIBRATING STRINGS.

431. The first general solution of the differential equation $\frac{\partial^2 y}{\partial t^2} = a^2 \frac{\partial^2 y}{\partial x^2}$, which determines the form of a string vibrating transversely, was given by d'Alembert* in the form $y = f(x + at) + \phi(x - at)$. He further shewed that, if $x = 0$, $x = l$, represent the fixed ends of the string, the form of the string at any time t is representable by $y = f(at + x) - f(at - x)$, where the function $f(z)$ is subject to the condition $f(z) = f(2l + z)$. D'Alembert was thus led to the search for analytical expressions which remain unaltered when $2l$ is added to the argument. In a second memoir, d'Alembert observed that the motion is determinate if the values of y and $\frac{\partial y}{\partial t}$ be assigned at some fixed time. Thus, in modern notation, if $y = f_1(x)$, $\frac{\partial y}{\partial t} = f_2(x)$, for $t = 0$, then for all values of x between 0 and l,

$$f(x) - f(-x) = f_1(x),$$
$$f(x) + f(-x) = \frac{1}{a}\int f_2(x)\,dx;$$

it follows that $f(x)$ is determined for all values of x between l and $-l$, and thence, by means of the condition $f(z) = f(2l + z)$, for all values of x.

The treatment of the same problem which was shortly afterwards given by Euler† was in form of a similar character to that of d'Alembert, but the difference of meaning assigned by these writers to the word "function" was of fundamental importance in the controversy which afterwards arose between the two mathematicians in relation to this problem. D'Alembert understood by a function $y = f(x)$, a single analytical expression, whereas Euler employed the same expression and notation to denote an arbitrarily given graph. Both, however, held the view that two analytical expressions which are equal for values of the variable in a given interval must also be equal for values of the variable outside that interval. D'Alembert argued that Euler's mode of determination of the function in the solution of the problem presupposes that y can be expressed in terms of x and t by means of a single analytical expression, and that thus an undue restriction is imposed upon the modes of vibration of the string. For example, in the case in which the initial figure of the string is polygonal, d'Alembert regarded the solution of the problem as impossible. The general effect of the controversy is to exhibit on the one hand the narrowness of the restriction of the conception of a function as held by d'Alembert, to functions

* *Memoirs of the Berlin Academy*, 1747, p. 214.
† *Memoirs of the Berlin Academy*, 1748, p. 69.

possessing at every point differential coefficients of all orders, and on the other hand the looseness of the conception of Euler that the ordinary methods of the Calculus are applicable without restriction to quite arbitrary functions.

432. The formal solution of the problem by means of trigonometrical series was given by Daniel Bernoulli* in a memoir in which he shewed that the differential equation and also the boundary conditions of the problem of the vibrating string, for the case in which there are no initial velocities, are formally satisfied by assuming

$$y = a_1 \sin \frac{\pi x}{l} \cos \frac{\pi a t}{l} + a_2 \sin \frac{2\pi x}{l} \cos \frac{2\pi a t}{l} + a_3 \sin \frac{3\pi x}{l} \cos \frac{3\pi a t}{l} + \dots.$$

He asserted that this represents the most general solution of the problem, and that the solutions of d'Alembert and Euler must therefore be contained in it. In a later memoir, he considered the case of a massless string loaded with n masses vibrating transversely, and indicated an indefinite increase in the number n. A criticism of Bernoulli's theory was published immediately afterwards by Euler, who pointed out that a consequence of Bernoulli's formula was that every arbitrarily assigned function of a variable x could be represented by a series of sines $a_1 \sin x + a_2 \sin 2x + a_3 \sin 3x + \dots$. This appeared to Euler to be a *reductio ad absurdum*, since such a series could represent only a function which is odd and periodic; the notion that a function could be capable of representation by a certain analytical expression only in a limited interval being contrary to established opinion at that time. Bernoulli's solution was consequently regarded by Euler as lacking in generality. A considerable controversy† took place on the subject between Bernoulli and d'Alembert.

This problem, together with the related problem of the propagation of plane waves in air, was next taken up by Lagrange‡, who obtained Euler's results by the method of starting with a finite number of masses fixed at intervals on a massless string, and then proceeding to the limit when the number of masses becomes indefinitely great. In the course of his analysis Lagrange came near to the determination of the form of the coefficients in the expansion of a function in a series of sines of multiples of the

* *Memoirs of the Berlin Academy*, 1753.

† For a detailed history of these controversies, see Burkhardt's *Bericht*, vol. I. The early history of the theory of trigonometrical series is given by Riemann in his memoir "Ueber die Darstellbarkeit einer Function durch eine trigonometrische Reihe," *Math. Werke*, p. 227. For the general history of the theory of these series see Sachs, "Versuch einer Geschichte der Darstellung willkürlicher Functionen einer Variabeln durch trigonometrische Reihen," *Schlömilch's Zeitschrift*, vol. XXV, supplement, and *Bulletin des sc. math.*, ser. 2, vol. IV, 1880; also Gibson "On the History of the Fourier Series," *Proceedings of the Edinburgh Math. Soc.*, vol. XI, p. 137.

‡ *Miscellanea Taurinensia*, vols. I, II, III.

argument. The defect of Lagrange's method lies in the lack of any investigation of the validity of the process of passing to the limit; no restrictions upon the nature of the arbitrary functions were recognized by him as necessary. The remarks made by Euler, d'Alembert and Bernoulli in the course of the discussion of Lagrange's work failed to elucidate the difficulties connected with this point, and no generally accepted theoretical views emerged from the lengthy controversies, the general course of which has been indicated.

The difficulties felt by the mathematicians of this period in regard to the generality of the representation of a function by a trigonometrical series arose in large measure from their restricted conception of the nature of a function. To them it was conceivable that a function given by a continuous curve might be so representable, but since they regarded a function obtained by piecing two or more such curves together, not as one function, but as several different functions, it seemed to them impossible that such a broken curve could be represented by one trigonometrical series; a separate series seemed to be required for each separate portion of the given composite curve. Moreover, the idea was unfamiliar that a particular mode of representation of a function need only be valid for some restricted range of values of the abscissa; and thus only a periodic curve was regarded as capable of being represented by means of a periodic series.

SPECIAL CASES OF TRIGONOMETRICAL SERIES.

433. Independently of the discussions of the problem of vibrating strings and of other physical problems, a number of trigonometrical series representing special functions of a simple character were obtained by Euler d'Alembert and Bernoulli. The methods employed by these writers for this purpose are of a character which fails to satisfy the requirements now regarded as necessary for the establishment of such results; moreover, in many cases the ranges of values of the variable for which the representations of the functions by the series are valid were not assigned.

For example, the series

$$\sin x - \tfrac{1}{2}\sin 2x + \tfrac{1}{3}\sin 3x - \tfrac{1}{4}\sin 4x + \dots,$$
$$\cos x - \tfrac{1}{4}\cos 2x + \tfrac{1}{9}\cos 3x - \tfrac{1}{16}\cos 4x - \dots,$$

were obtained by Euler*, as representing $\frac{1}{2}x$, $\frac{1}{12}\pi^2 - \frac{1}{4}x^2$ respectively; the range of values of x $(-\pi, \pi)$ for which these representations are valid was however not given by Euler, who appeared to regard them as valid for all values of x. These series were obtained by integration of the series

* *Petrop. N. Comm.* 1754–55, and *Petrop. N. Acta*, 1789.

$\cos x + \cos 2x + \cos 3x + \dots$, the sum of which was maintained by Euler to be $-\frac{1}{2}$.

By D. Bernoulli* the series $\sum_{n=1}^{\infty} \frac{1}{n} \sin nx$ was obtained as a representation of $\frac{1}{2}(\pi - x)$, and the range of values of $x\,(0, 2\pi)$ for which this representation is valid was assigned. It was also observed that the sum of the series is discontinuous for $x = 0, 2\pi, 4\pi, \dots$. The following series were also obtained by Bernoulli, and the ranges of the validity of the equations were assigned:—

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \cos nx = \frac{1}{6}\pi^2 - \frac{1}{2}\pi x + \frac{1}{4}x^2,$$

$$\sum_{n=1}^{\infty} \frac{1}{n^3} \sin nx = \frac{1}{6}\pi^2 x - \frac{1}{4}\pi x^2 + \frac{1}{12}x^3,$$

$$\sum_{n=1}^{\infty} \frac{1}{n^4} \cos nx = \frac{1}{90}\pi^4 - \frac{1}{12}\pi^2 x^2 + \frac{1}{12}\pi x^3 - \frac{1}{48}x^4,$$

$$\sum_{n=1}^{\infty} \frac{1}{n^5} \sin nx = \frac{1}{90}\pi^4 x - \frac{1}{36}\pi^2 x^3 + \frac{1}{48}\pi x^4 - \frac{1}{240}x^5,$$

$$\sum_{n=1}^{\infty} \frac{1}{n} \cos nx = \frac{1}{2}\log\frac{1}{2(1-\cos x)}.$$

It was remarked by Bernoulli that the sums of these series have discontinuities at $x = 0, 2\pi, 4\pi, \dots$.

The following results among others obtained by Euler may here be mentioned:—

$$\frac{1}{4}\pi = \sum_{r=0}^{\infty} (-1)^r \frac{\cos(2r+1)x}{2r+1},$$

$$\frac{1}{4}\pi x = \sum_{r=0}^{\infty} (-1)^r \frac{\sin(2r+1)x}{(2r+1)^2},$$

$$\frac{1}{8}\pi\left(\frac{1}{4}\pi^2 - x^2\right) = \sum_{r=0}^{\infty} (-1)^r \frac{\cos(2r+1)x}{(2r+1)^3}.$$

The true range of validity of these equations will be given later.

LATER HISTORY OF THE THEORY.

434. No further advance was made in the subject until 1807, when Fourier, in a memoir on the Theory of Heat presented† to the French Academy, laid down the proposition that an arbitrary function given graphically by means of a curve, which may be broken by (ordinary) discontinuities, is capable of representation by means of a single trigonometrical series. This theorem is said to have been received by Lagrange with astonishment and incredulity.

* *Petrop. N. Comm.* 1772.

† *Bulletin des sciences de la soc. philomathique*, vol. I, p. 122.

Fourier shewed, in a variety of special cases, that a function $f(x)$ is representable for values of x between $-\pi$ and π, by the series

$$\tfrac{1}{2}a_0 + (a_1 \cos x + b_1 \sin x) + (a_2 \cos 2x + b_2 \sin 2x) + \ldots$$

where $$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x) \cos nx\, dx, \quad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x) \sin nx\, dx,$$

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\, dx.$$

Fourier's results in connection with this subject are best studied in the collected form in which they appear in his *Théorie de la Chaleur*, published in 1822. Trigonometrical series of the above form, in which the coefficients are determined as above, are known as Fourier's series. It should, however, be remarked that Fourier also studied other trigonometrical series, in which the cosines and sines do not proceed by integral multiples of the argument. These latter series will not be considered in this work.

Although Fourier attained to correct views as to the nature of the convergence of the infinite series he employed, he did not give any complete general proof that the series in the general case actually converges to the value of the function; he indicates* however on general lines a process of verification of such convergence which was not actually carried out until Dirichlet took up the subject.

435. An attempt to prove Fourier's theorem was made by Poisson, who started with the formula†

$$\int_{-\pi}^{\pi} f(x') \frac{1-h^2}{1-2h\cos(x-x')+h^2}\, dx' = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x')\, dx' + \frac{1}{\pi}\sum_{n=1}^{\infty} h^n \int_{-\pi}^{\pi} f(x') \cos n(x-x')\, dx',$$

which holds provided $-1 < h < 1$.

Poisson proceeded to shew that as h approaches the limit 1, the integral on the left-hand side of the equation approaches the limit $f(x)$, and argued that $f(x)$ is represented by the series obtained by putting $h = 1$, on the right-hand side. Apart from the questions connected with the limit of the integral on the left-hand side, the conclusion is invalid unless it is shewn that the series obtained by putting $h=1$, is convergent. In accordance with a known theorem, given by Abel, for power series (see § 356), in case the power series is convergent for $h = 1$, it converges to the limit of the sum of the series for values of h which are < 1, as h approaches the value 1; but no conclusion can be made as to whether the series is really convergent,

* See the *Théorie de la chaleur*, chap. IX, especially § 423.

† *Journ. de l'école polyt.* cah. 19, 1823, p. 404. See also his *Théorie analytique de la chaleur*.

or not, when $h=1$. A direct investigation of its convergence would be required to make the proof a valid one.

Two proofs of the validity of the representation were given by Cauchy; one at least of these is certainly invalid in its original form. Both of them depend upon the theory of functions of a complex variable, and will consequently not be discussed here. An example of an invalid proof of a similar character to one of Cauchy's and also to Poisson's is the proof given in Thomson and Tait's *Natural Philosophy*.

In 1829, Dirichlet* gave a proof that, in an extensive class of cases, Fourier's series actually converges to the value of the function. His proof, the first rigid one, was based upon a recognition of the distinction between absolutely convergent, and conditionally convergent, series. Since a Fourier's series, when convergent, is not necessarily absolutely convergent, it is impossible to obtain a proof of the convergence from the law according to which the terms diminish, as Cauchy had attempted to do. As Dirichlet's proof, apart from its historical interest, still repays a careful study on account of the light it throws upon the mode of convergence of the series, it will be given below, with some modifications and extensions which arise from later advances in the Theory of Functions.

THE FORMAL EXPRESSION OF FOURIER'S SERIES.

436. Let $f(x)$ denote a limited function, defined for the interval $(0, l)$ of the variable x. A finite trigonometrical series of the form

$$a_1 \sin \frac{\pi x}{l} + a_2 \sin \frac{2\pi x}{l} + \ldots + a_s \sin \frac{s\pi x}{l} + \ldots + a_{n-1} \sin \frac{(n-1)\pi x}{l}$$

can be so determined that its value is equal to that of the function $f(x)$ at each of the points $x = \frac{l}{n}, \frac{2l}{n}, \frac{3l}{n}, \ldots \frac{\overline{n-1}\,l}{n}$. It must be shewn that the coefficients $a_1, a_2 \ldots a_{n-1}$ can be determined by means of the linear equations

$$f\left(\frac{l}{n}\right) = a_1 \sin \frac{\pi}{n} + a_2 \sin \frac{2\pi}{n} + \ldots + a_{n-1} \sin \frac{(n-1)\pi}{n},$$

$$f\left(\frac{2l}{n}\right) = a_1 \sin \frac{2\pi}{n} + a_2 \sin \frac{2 \cdot 2\pi}{n} + \ldots + a_{n-1} \sin \frac{2(n-1)\pi}{n},$$

....

* *Crelle's Journal*, vol. IV, "Sur la convergence des séries trigonométriques, qui servent à représenter une fonction arbitraire entre des limites données." See also his memoir in Dove and Moser's *Repertorium für Physik*, vol. I, 1837. Papers by Dircksen, *Crelle's Journal*, vol. IV, and by Bessel, *Astron. Nachrichten*, vol. XVI, are on similar lines to those of Dirichlet, but of inferior importance.

....

$$f\left(\frac{rl}{n}\right) = a_1 \sin\frac{r\pi}{n} + a_2 \sin\frac{2r\pi}{n} + \ldots + a_{n-1} \sin\frac{r(n-1)\pi}{n},$$

....

$$f\left(\frac{\overline{n-1}\,l}{n}\right) = a_1 \sin\frac{(n-1)\pi}{n} + a_2 \sin\frac{2(n-1)\pi}{n} + \ldots + a_{n-1} \sin\frac{(n-1)(n-1)\pi}{n}.$$

Multiply the expressions on the two sides of these equations by

$$\sin\frac{s\pi}{n},\ \sin\frac{2s\pi}{n}, \ldots \sin\frac{(n-1)s\pi}{n}$$

respectively, and add the expressions on each side together. It can easily be verified that

$$\sin\frac{r\pi}{n}\sin\frac{s\pi}{n} + \sin\frac{2r\pi}{n}\sin\frac{2s\pi}{n} + \ldots + \sin\frac{(n-1)r\pi}{n}\sin\frac{(n-1)s\pi}{n} = 0,$$

provided r and s are unequal integers not greater than $n-1$; and also it can be shewn that

$$\sin^2\frac{s\pi}{n} + \sin^2\frac{2s\pi}{n} + \ldots + \sin^2\frac{(n-1)s\pi}{n} = \frac{1}{2}n.$$

Using these two identities, we have at once

$$a_s = \frac{2}{n}\left[f\left(\frac{l}{n}\right)\sin\frac{s\pi}{n} + f\left(\frac{2l}{n}\right)\sin\frac{2s\pi}{n} + \ldots + f\left(\frac{\overline{n-1}\,\pi}{n}\right)\sin\frac{s(n-1)\pi}{n}\right];$$

and thus the coefficients in the series have been determined so that the series satisfies the prescribed condition. Let us now assume that the function $f(x)$ is integrable in accordance with Riemann's definition, and let the number n be indefinitely increased. The limit of the expression for a_s is then seen to be $\frac{2}{l}\int_0^l f(x')\sin\frac{s\pi x'}{l}\,dx'$. This process suggests the possibility that the function $f(x)$ may be represented by the infinite series

$$a_1 \sin\frac{\pi x}{l} + a_2 \sin\frac{2\pi x}{l} + \ldots + a_s \sin\frac{s\pi x}{l} + \ldots$$

where the coefficients a_s are given by

$$a_s = \frac{2}{l}\int_0^l f(x')\sin\frac{s\pi x'}{l}\,dx,$$

for points x within the interval $(0, l)$. It will be observed that the series cannot possibly represent the function at the point $x=0$, unless $f(0)=0$; nor at the point $x=l$, unless $f(l)=0$. This limiting process is entirely insufficient to shew either that the infinite series converges at all, or that, when it does converge, its limiting sum is at any point equal to the value of the function $f(x)$ at that point.

It will later be shewn by various methods that, for extensive classes of functions, the series

$$\frac{2}{l}\sum_{s=1}^{\infty}\sin\frac{s\pi x}{l}\int_0^l f(x')\sin\frac{s\pi x'}{l}dx' \quad\text{.....................(1)}$$

actually converges to the value $f(x)$, for values of x within the interval $(0, l)$, at which $f(x)$ is continuous. This series is known as *Fourier's sine series.*

Let us now assume that the function $f(x)\sin\frac{\pi x}{l}$ is represented within the interval $(0, l)$ by the Fourier's sine series.

This series is, in the present case, of the form

$$\frac{2}{l}\sum_{s=1}^{\infty}\sin\frac{s\pi x}{l}\int_0^l f(x')\sin\frac{\pi x'}{l}\sin\frac{s\pi x'}{l}dx',$$

which is equivalent to

$$\frac{1}{l}\sum_{s=1}^{\infty}\sin\frac{s\pi x}{l}\int_0^l f(x')\left[\cos\frac{\overline{s-1}\,\pi x'}{l}-\cos\frac{\overline{s+1}\,\pi x'}{l}\right]dx',$$

or to

$$\frac{1}{l}\sin\frac{\pi x}{l}\int_0^l f(x')\,dx'+\frac{1}{l}\sum_{s=1}^{\infty}\left\{\sin\frac{\overline{s+1}\,\pi x}{l}-\sin\frac{\overline{s-1}\,\pi x}{l}\right\}\int_0^l f(x')\cos\frac{s\pi x'}{l}dx';$$

and this by hypothesis represents the function $f(x)\sin\frac{\pi x}{l}$.

It thus appears that, on the assumptions made, the function $f(x)$ is represented by the series

$$\frac{1}{l}\int_0^l f(x')\,dx'+\frac{2}{l}\sum_{s=1}^{\infty}\cos\frac{s\pi x}{l}\int_0^l f(x')\cos\frac{s\pi x'}{l}dx' \quad\text{...........(2).}$$

This series (2) is of the form

$$\beta_0+\beta_1\cos\frac{\pi x}{l}+\beta_2\cos\frac{2\pi x}{l}+\ldots+\beta_s\cos\frac{s\pi x}{l}+\ldots,$$

and is known as *Fourier's cosine series.*

The cosine series, unlike the sine series, may possibly converge to the values $f(0)$, $f(l)$, for $x=0$, l respectively, when these functional values are not necessarily zero.

437. Assuming for the present that the function $f(x)$ may be represented for the points of the interval $(0, l)$ by either of these series (1) and (2), we proceed to consider some obvious properties of the series themselves. The sum of the sine series (1) has, for the point $-x$, the same value, with the opposite sign, as for the point x. If then we suppose that the function $f(x)$ is defined not only for the interval $(0, l)$, but for the interval $(-l, l)$, it appears that the series can represent the function for the whole interval $(-l, l)$, only in case $f(-x)=-f(x)$; that is, in case the function $f(x)$ be odd.

Further, the series (1) is unaltered by adding to x any multiple of $2l$, and thus the series, considered as existent for all values of x, defines a periodic function, of period $2l$. If $f(x)$ be defined for all values of x, it can only be represented by the series, for all such values of x, provided $f(x)$ is periodic and of period $2l$, and also $f(x) = -f(-x)$; otherwise the representation of the function by the series is valid only for the interval $(0, l)$.

The cosine series (2) is unaltered by changing x into $-x$; therefore the series represents the function $f(x)$ for the interval $(-l, l)$, only when $f(-x) = f(x)$, *i.e.* when $f(x)$ is an even function. The cosine series, like the sine series, considered as existent for all values of x, is periodic, and of period $2l$; therefore the series can represent a function $f(x)$, defined for all values of x, only when $f(x)$ is periodic with period $2l$, and also $f(x) = f(-x)$.

It is thus seen that, if the function $f(x)$ be defined for the interval $(-l, l)$, it is in general not represented by either the sine or the cosine series for the whole of that interval, although it may be represented by both the series for the interval $(0, l)$. For the part of the function $f(x)$ in the interval $(-l, 0)$ is in general independent of the part in the interval $(0, l)$; neither of the relations $f(-x) = -f(x)$, $f(-x) = f(x)$ being in general satisfied. In fact there is in general no relation between the values of a function, defined for the interval $(-l, l)$, at the two points $-x$, x.

It is however possible to obtain, from the series (1) and (2), a series containing both sines and cosines, such as to represent the function $f(x)$ for the whole interval $(-l, l)$. The function $\frac{1}{2}\{f(x) + f(-x)\}$ is an even function, defined for the whole interval $(-l, l)$, and in accordance with the assumptions, representable for that interval by the series

$$\frac{1}{2l}\int_0^l \{f(x') + f(-x')\}\,dx' + \frac{1}{l}\sum_{s=1}^{\infty} \cos\frac{s\pi x}{l}\int_0^l [f(x') + f(-x')]\cos\frac{s\pi x'}{l}\,dx'.$$

Again, the function $\frac{1}{2}\{f(x) - f(-x)\}$ is an odd function, defined for the whole interval $(-l, l)$, and is accordingly representable by

$$\frac{1}{l}\sum_{s=1}^{\infty} \sin\frac{s\pi x}{l}\int_0^l [f(x') - f(-x')]\sin\frac{s\pi x'}{l}\,dx'.$$

By addition of the two series, we find the series

$$\frac{1}{2l}\int_{-l}^l f(x')\,dx' + \frac{1}{l}\sum_{s=1}^{\infty}\int_{-l}^l \cos\frac{s\pi}{l}(x - x')f(x')\,dx' \;\ldots\ldots\ldots\ldots(3),$$

which is of the form

$$\tfrac{1}{2}\alpha_0 + \left(\alpha_1 \cos\frac{\pi x}{l} + \beta_1 \sin\frac{\pi x}{l}\right) + \left(\alpha_2 \cos\frac{2\pi x}{l} + \beta_2 \sin\frac{2\pi x}{l}\right) + \ldots,$$

as representing the function $f(x)$ for the interval $(-l, l)$. This series (3) is known as *Fourier's series*, the sine and cosine series being regarded as the particular cases of it which arise when $f(-x) = -f(x)$, or $f(-x) = f(x)$ respectively.

438. With certain assumptions, the form of the series (3) may be obtained directly. Let it be assumed that a function $f(x)$, defined for the interval $(-l, l)$, can be represented by the series

$$\tfrac{1}{2}\alpha_0 + \left(\alpha_1 \cos\frac{\pi x}{l} + \beta_1 \sin\frac{\pi x}{l}\right) + \ldots + \left(\alpha_n \cos\frac{n\pi x}{l} + \beta_n \sin\frac{n\pi x}{l}\right) + \ldots,$$

in the sense that this series converges to $f(x)$ for each value of x in the interval. If it be further assumed that the convergence of the series is uniform in the interval, and thus that $f(x)$ is continuous and consequently integrable in the interval $(-l, l)$, we may submit the series to a term by term integration, even when it is multiplied by $\cos\frac{n\pi x}{l}$, or by $\sin\frac{n\pi x}{l}$. It would be sufficient for our purpose to assume that the series, without being necessarily uniformly convergent, is still such that a term by term integration is admissible, in accordance with the criteria investigated in § 383, and that $f(x)$ is integrable in $(-l, l)$.

Making use of the fundamental property of circular functions, represented by the formula $\int_{-l}^{l} \begin{matrix}\cos\\ \sin\end{matrix}\frac{n\pi x}{l} \begin{matrix}\cos\\ \sin\end{matrix}\frac{n'\pi x}{l}\,dx = 0$, where n and n' are any unequal integers, and observing that $\int_{-l}^{l} \cos^2\frac{n\pi x}{l}\,dx = \int_{-l}^{l} \sin^2\frac{n\pi x}{l}\,dx = l$, we thus find that

$$\alpha_0 = \frac{1}{l}\int_{-l}^{l} f(x')\,dx', \quad \alpha_n = \frac{1}{l}\int_{-l}^{l} f(x')\cos\frac{n\pi x'}{l}\,dx', \quad \beta_n = \frac{1}{l}\int_{-l}^{l} f(x')\sin\frac{n\pi x'}{l}\,dx'.$$

Therefore we have, for the interval $(-l, l)$, as the series representing $f(x)$,

$$\frac{1}{2l}\int_{-l}^{l} f(x')\,dx' + \sum_{n=1}^{\infty}\left\{\frac{1}{l}\cos\frac{n\pi x}{l}\int_{-l}^{l} f(x')\cos\frac{n\pi x'}{l}\,dx' \right.$$
$$\left. + \frac{1}{l}\sin\frac{n\pi x}{l}\int_{-l}^{l} f(x')\sin\frac{n\pi x'}{l}\,dx'\right\},$$

or

$$\frac{1}{2l}\int_{-l}^{l} f(x')\,dx' + \sum_{n=1}^{\infty}\frac{1}{l}\int_{-l}^{l} f(x')\cos\frac{n\pi}{l}(x - x')\,dx'.$$

If we replace $\frac{\pi x}{l}$ by x, no essential difference will be made in the formula; thus there is no loss of generality in taking the interval $(-\pi, \pi)$ to be the interval in which $f(x)$ is defined, and for which it is represented by

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x')\,dx' + \sum_{n=1}^{\infty}\frac{1}{\pi}\int_{-\pi}^{\pi} f(x')\cos n(x - x')\,dx' \quad \ldots\ldots\ldots(4).$$

This expression (4) will be taken as the standard form of Fourier's series.

439. The form of the series (4) having now been obtained by purely tentative processes, the reverse course will be adopted of taking the series itself as the starting-point, and subjecting it to an examination with a view of discovering under what circumstances it is convergent for all or some of the values of x in the interval $(-\pi, \pi)$, and of determining the value to which it converges in case it is convergent.

In order that the series (4) may exist, whether it converge or not, it is necessary that the coefficients $\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x')\,dx'$, $\frac{1}{\pi}\int_{-\pi}^{\pi} f(x')\cos nx'dx'$, $\frac{1}{\pi}\int_{-\pi}^{\pi} f(x')\sin nx'dx'$ should have definite meanings. Until quite recently it has consequently been assumed that $f(x)$ is either limited in the interval $(-\pi, \pi)$ and integrable in accordance with the definition of Riemann, or else that $f(x)$ is unlimited in that interval, but possesses an improper integral in accordance with one of the definitions which have been given of improper integrals. The recent extension of the definition of integration, by Lebesgue, to the case of functions which are not necessarily integrable in accordance with Riemann's definition leads to a corresponding extension of the domain of Fourier's series. It has been proposed by Lebesgue* to assign to the series (4) the name *Fourier's series*, in every case in which $f(x)$ is a summable function in the interval $(-\pi, \pi)$, whether the summable function be limited in the interval or not, provided that, when the function is unlimited, it be still integrable. Since only those improper integrals which are absolutely convergent are included in Lebesgue's definition, there remains the case in which the coefficients exist only as non-absolutely convergent integrals; in this case Lebesgue has proposed to name the series *generalized Fourier's series*. This terminology will be here adopted.

440. It being assumed that $f(x)$, as defined for the interval $(-\pi, \pi)$, is such that the coefficients in the series (4) have definite meanings, it is easy to express the sum of a finite number of terms of the series as a definite integral.

Since $\frac{1}{2} + \cos\theta + \cos 2\theta + \ldots + \cos n\theta = \dfrac{\sin\frac{1}{2}(2n+1)\theta}{2\sin\frac{1}{2}\theta}$, we see that S_{2n+1}, the sum of the first $2n+1$ terms of the series

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x')\,dx' + \sum_{n=1}\left\{\frac{1}{\pi}\cos nx\int_{-\pi}^{\pi} f(x')\cos nx'dx' + \frac{1}{\pi}\sin nx\int_{-\pi}^{\pi} f(x')\sin nx'dx'\right\},$$

* Lebesgue's treatment of the series is contained in a memoir "Sur les séries trigonométriques," *Annales sc. de l'École normale, supérieure*, ser. 3, vol. xx, 1903; in a memoir "Sur la convergence des Séries de Fourier," *Math. Annalen*, vol. LXIV, 1905; and in the *Leçons sur les séries trigonométriques*, 1906.

is given by
$$S_{2n+1}=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x')\frac{\sin(2n+1)\dfrac{x'-x}{2}}{2\sin\dfrac{x'-x}{2}}\,dx'.$$

If we change the independent variable in this integral from x' to z, where $x'=x+2z$, and write $2n+1=m$, the expression becomes
$$S_m=\frac{1}{\pi}\int_{-\frac{1}{2}(\pi+x)}^{\frac{1}{2}(\pi-x)}f(x+2z)\frac{\sin mz}{\sin z}\,dz.$$

In order that the series may converge at a point x, it is necessary that S_m should converge to a definite limit, as the odd integer m is indefinitely increased.

It was first shewn by Dirichlet that, for an important class of functions $f(x)$, S_m converges to the value $f(x)$ at every point x in the interior of the interval $(-\pi, \pi)$ at which $f(x)$ is continuous; that, at a point of ordinary discontinuity of $f(x)$, it converges to the value $\frac{1}{2}\{f(x+0)+f(x-0)\}$, which is not of course necessarily equal to $f(x)$; and that at the points $x=\pi$ or $-\pi$, it converges to the value $\frac{1}{2}\{f(\pi-0)+f(-\pi+0)\}$.

An account will be given in the present Chapter of the investigations, by various writers, which have as their object the determination of sufficient conditions to be satisfied by the function $f(x)$ in order that the series may converge either throughout the whole interval, or at particular points of that interval. It will appear that the convergence or non-convergence of the series, at a particular point x, really depends only upon the nature of the function in an arbitrarily small neighbourhood of that point, and is independent of the general character of the function throughout the interval; this general character being limited only by the necessity for the existence of the coefficients of the series. These investigations have resulted in the discovery of sufficient conditions of considerable width, which suffice for the convergence of the series either at particular points, or generally throughout the interval for which the function is defined. The necessary and sufficient conditions for the convergence of the series at a point of the interval, or throughout any portion of the interval, have not been obtained. This is not surprising, in view of the very general character of the problem; and indeed it is not improbable that no such necessary and sufficient conditions may be obtainable. It is possible that the mere fact of the convergence of the series at a point characterizes the nature of the function in the neighbourhood of that point in a manner incapable of reduction to any other form; so that although the characteristics of various sub-classes of the functions satisfying this condition may be obtained, as has in fact been done, yet the whole class of such functions has no property capable of being stated in any form different from the mere statement of the fact of the convergence of

439. The form of the series (4) having now been obtained by purely tentative processes, the reverse course will be adopted of taking the series itself as the starting-point, and subjecting it to an examination with a view of discovering under what circumstances it is convergent for all or some of the values of x in the interval $(-\pi, \pi)$, and of determining the value to which it converges in case it is convergent.

In order that the series (4) may exist, whether it converge or not, it is necessary that the coefficients $\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x')\,dx'$, $\frac{1}{\pi}\int_{-\pi}^{\pi} f(x')\cos nx'dx'$, $\frac{1}{\pi}\int_{-\pi}^{\pi} f(x')\sin nx'dx'$ should have definite meanings. Until quite recently it has consequently been assumed that $f(x)$ is either limited in the interval $(-\pi, \pi)$ and integrable in accordance with the definition of Riemann, or else that $f(x)$ is unlimited in that interval, but possesses an improper integral in accordance with one of the definitions which have been given of improper integrals. The recent extension of the definition of integration, by Lebesgue, to the case of functions which are not necessarily integrable in accordance with Riemann's definition leads to a corresponding extension of the domain of Fourier's series. It has been proposed by Lebesgue* to assign to the series (4) the name *Fourier's series*, in every case in which $f(x)$ is a summable function in the interval $(-\pi, \pi)$, whether the summable function be limited in the interval or not, provided that, when the function is unlimited, it be still integrable. Since only those improper integrals which are absolutely convergent are included in Lebesgue's definition, there remains the case in which the coefficients exist only as non-absolutely convergent integrals; in this case Lebesgue has proposed to name the series *generalized Fourier's series*. This terminology will be here adopted.

440. It being assumed that $f(x)$, as defined for the interval $(-\pi, \pi)$, is such that the coefficients in the series (4) have definite meanings, it is easy to express the sum of a finite number of terms of the series as a definite integral.

Since $\frac{1}{2} + \cos\theta + \cos 2\theta + \ldots + \cos n\theta = \frac{\sin\frac{1}{2}(2n+1)\theta}{2\sin\frac{1}{2}\theta}$, we see that S_{2n+1}, the sum of the first $2n+1$ terms of the series

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x')\,dx' + \sum_{n=1}\left\{\frac{1}{\pi}\cos nx\int_{-\pi}^{\pi} f(x')\cos nx'dx' + \frac{1}{\pi}\sin nx\int_{-\pi}^{\pi} f(x')\sin nx'dx'\right\},$$

* Lebesgue's treatment of the series is contained in a memoir "Sur les séries trigonométriques," *Annales sc. de l'école normale, supérieure*, ser. 3, vol. XX, 1903; in a memoir "Sur la convergence des Séries de Fourier," *Math. Annalen*, vol. LXIV, 1905; and in the *Leçons sur les séries trigonométriques*, 1906.

is given by $$S_{2n+1} = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x') \frac{\sin(2n+1)\frac{x'-x}{2}}{2\sin\frac{x'-x}{2}} dx'.$$

If we change the independent variable in this integral from x' to z, where $x' = x + 2z$, and write $2n + 1 = m$, the expression becomes

$$S_m = \frac{1}{\pi}\int_{-\frac{1}{2}(\pi+x)}^{\frac{1}{2}(\pi-x)} f(x+2z)\frac{\sin mz}{\sin z} dz.$$

In order that the series may converge at a point x, it is necessary that S_m should converge to a definite limit, as the odd integer m is indefinitely increased.

It was first shewn by Dirichlet that, for an important class of functions $f(x)$, S_m converges to the value $f(x)$ at every point x in the interior of the interval $(-\pi, \pi)$ at which $f(x)$ is continuous; that, at a point of ordinary discontinuity of $f(x)$, it converges to the value $\frac{1}{2}\{f(x+0) + f(x-0)\}$, which is not of course necessarily equal to $f(x)$; and that at the points $x = \pi$ or $-\pi$, it converges to the value $\frac{1}{2}\{f(\pi - 0) + f(-\pi + 0)\}$.

An account will be given in the present Chapter of the investigations, by various writers, which have as their object the determination of sufficient conditions to be satisfied by the function $f(x)$ in order that the series may converge either throughout the whole interval, or at particular points of that interval. It will appear that the convergence or non-convergence of the series, at a particular point x, really depends only upon the nature of the function in an arbitrarily small neighbourhood of that point, and is independent of the general character of the function throughout the interval; this general character being limited only by the necessity for the existence of the coefficients of the series. These investigations have resulted in the discovery of sufficient conditions of considerable width, which suffice for the convergence of the series either at particular points, or generally throughout the interval for which the function is defined. The necessary and sufficient conditions for the convergence of the series at a point of the interval, or throughout any portion of the interval, have not been obtained. This is not surprising, in view of the very general character of the problem; and indeed it is not improbable that no such necessary and sufficient conditions may be obtainable. It is possible that the mere fact of the convergence of the series at a point characterizes the nature of the function in the neighbourhood of that point in a manner incapable of reduction to any other form; so that although the characteristics of various sub-classes of the functions satisfying this condition may be obtained, as has in fact been done, yet the whole class of such functions has no property capable of being stated in any form different from the mere statement of the fact of the convergence of the series. It will appear that there exist functions, even continuous functions, for which the series fails to converge at every point of the interval belonging to an everywhere-dense set of points.

Recent investigations, an account of which will be given, shew that the coefficients of Fourier's series have important properties which are related to the functional values, independently of whether the series converges or not; so that a divergent Fourier's series may be employed, in accordance with recent ideas, for the representation of the function in a certain sense.

PARTICULAR CASES OF FOURIER'S SERIES.

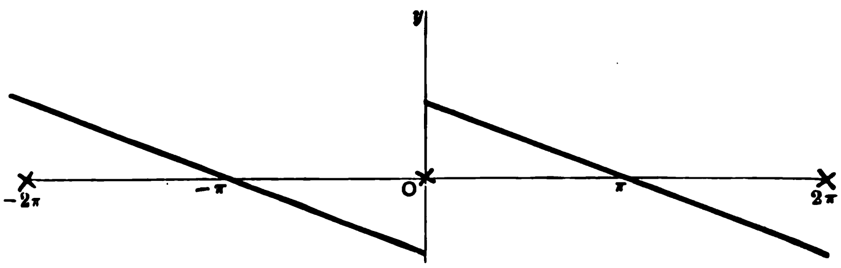
441. Before proceeding to the theoretical investigations relating to the convergence and the properties of Fourier's series, it will be instructive to consider some simple cases of the use of the series. It will be assumed that, for the functions employed, the series corresponding to a function $f(x)$ converges at every point to the value $\frac{1}{2}\{f(x+0)+f(x-0)\}$.

If we employ the sine series to represent the function defined, for the interval $(0, \pi)$, by $y=\frac{1}{2}(\pi-x)$, we find on evaluation that

$$\frac{2}{\pi}\int_0^\pi \tfrac{1}{2}(\pi-x)\sin nx dx = \frac{1}{n};$$

and thus the series is of the form

$$\sin x + \tfrac{1}{2}\sin 2x + \tfrac{1}{3}\sin 3x + \ldots + \frac{1}{n}\sin nx + \ldots.$$



The function defined for all values of x by

$$y = \sin x + \tfrac{1}{2}\sin 2x + \ldots + \frac{1}{n}\sin nx + \ldots,$$

is represented graphically in the figure. The function is discontinuous at the points $0, 2\pi, 4\pi, \ldots -2\pi, -4\pi, \ldots$; the functional value being zero at all those points. It is seen that the series represents the function $\frac{1}{2}(\pi-x)$, not only for the interval $(0, \pi)$, but for the interval $(0, 2\pi)$, except at the points $x=0$, $x=2\pi$, where the sum of the series is zero. For the interval $(-2\pi, 0)$

the function represented by the series is $-\frac{1}{2}(\pi + x)$, except at the ends of the interval.

This series may be employed to illustrate some important points connected with the convergence of the series in the neighbourhood of the point $x = 0$, at which the function represented by the series is discontinuous. To this end we shall examine the series by a method employed by Fourier*, and further developed by Kneser†.

Denoting $\sin x + \frac{1}{2}\sin 2x + \ldots + \frac{1}{n}\sin nx$, by $s_n(x)$, we have

$$\frac{ds_n(x)}{dx} = \cos x + \cos 2x + \ldots + \cos nx = \frac{\sin(n+\frac{1}{2})x}{2\sin\frac{1}{2}x} - \frac{1}{2};$$

therefore

$$\begin{aligned} s_n(x) &= \frac{1}{2}\int_0^x \frac{\sin(n+\frac{1}{2})x}{\sin\frac{1}{2}x}dx - \frac{1}{2}x \\ &= \int_0^x \frac{\sin(n+\frac{1}{2})x}{x}dx - \frac{1}{2}x + \int_0^x \sin(n+\frac{1}{2})x\frac{(x-2\sin\frac{1}{2}x)}{2x\sin\frac{1}{2}x}dx \\ &= \int_0^{(n+\frac{1}{2})x}\frac{\sin z}{z}dz - \frac{1}{2}x + I(x). \end{aligned}$$

On integrating by parts, we find that

$$I(x) = -\frac{x-2\sin\frac{1}{2}x}{2x\sin\frac{1}{2}x}\frac{\cos(n+\frac{1}{2})x}{n+\frac{1}{2}} + \int_0^x \frac{\cos(n+\frac{1}{2})x}{n+\frac{1}{2}}\cdot\frac{4\sin^2\frac{1}{2}x - x^2\cos\frac{1}{2}x}{4x^2\sin^2\frac{1}{2}x}dx.$$

The expressions $\dfrac{x-2\sin\frac{1}{2}x}{2x\sin\frac{1}{2}x}$, $\dfrac{4\sin^2\frac{1}{2}x - x^2\cos\frac{1}{2}x}{4x^2\sin^2\frac{1}{2}x}$ both become indefinitely great, as x increases up to 2π; but if x be confined to the interval $(0, b)$ where $0 < b < 2\pi$, they are both limited functions. It follows, since

$$|\cos(n+\tfrac{1}{2}x)| \leqq 1,$$

that a positive number A can be determined, independent of n and x, such that $|I(x)| < A/(n+\frac{1}{2})$, provided x is in the interval $(0, b)$. Hence it appears that $I(x)$ has the limit zero, when n is indefinitely increased, whether x varies with n or not; in fact $|I(x)|$ is arbitrarily small for sufficiently great values of n.

We have now

$$s_n(x) - s(x) = \int_0^{(n+\frac{1}{2})x}\frac{\sin z}{z}dz - \frac{1}{2}\pi + \frac{\theta A}{n+\frac{1}{2}},$$

provided $0 \leqq x \leqq b$; where θ is such that $-1 < \theta < 1$.

Also
$$\frac{d}{dx}\{s_n(x) - s(x)\} = \frac{1}{2}\frac{\sin(n+\frac{1}{2})x}{\sin\frac{1}{2}x}, \text{ if } 0 < x \leqq b;$$

* *Théorie de la chaleur*, chap. III, § 3.

† *Grunert's Archiv*, ser. 3, vol. VII, 1904. See also Bôcher's "Introduction to the theory of Fourier's series," *Annals of Mathematics*, ser. 2, vol. VII, where numerical details are worked out.

and therefore $s_n(x) - s(x)$ has maxima and minima at the points $x = \frac{\lambda\pi}{n+\frac{1}{2}}$, where $\lambda = 1, 2, 3, \ldots$.

It can now be shewn that, for sufficiently large values of n, at least,

$$s_n\left(\frac{2\pi}{2n+1}\right) - s\left(\frac{2\pi}{2n+1}\right),\ s_n\left(\frac{4\pi}{2n+1}\right) - s\left(\frac{4\pi}{2n+1}\right),\ s_n\left(\frac{6\pi}{2n+1}\right) - s\left(\frac{6\pi}{2n+1}\right), \ldots$$

are alternately positive and negative, the first of these differences being positive.

We have $\int_0^{\lambda\pi} \frac{\sin z}{z}\,dz = \int_0^{\pi} \sin z \left(\frac{1}{z} - \frac{1}{z+\pi} + \frac{1}{z+2\pi} - \ldots + \frac{(-1)^{\lambda+1}}{z+\overline{\lambda-1}\,\pi}\right) dz$

$$= u_1 - u_2 + u_3 - \ldots + (-1)^{\lambda+1} u_\lambda,$$

where $u_1, u_2, \ldots u_\lambda$ are all positive, and $u_1 > u_2 > u_3 \ldots > u_\lambda$. Also

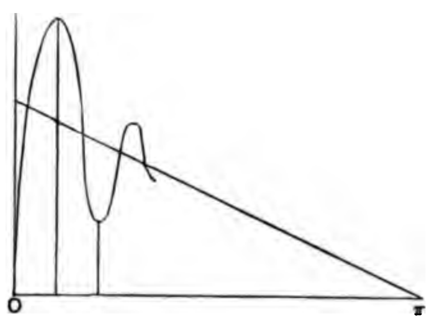
$$u_\lambda < \frac{1}{\overline{\lambda-1}\,\pi}\int_0^{\pi} \sin z\,dz < \frac{2}{\overline{\lambda-1}\,\pi};$$

hence $\lim_{\lambda=\infty} u_\lambda = 0$.

Further, it is well known that $\lim_{\lambda=\infty}\int_0^{\lambda\pi} \frac{\sin z}{z}\,dz$, which is the improper integral $\int_0^{\infty} \frac{\sin z}{z}\,dz$, is equal to $\frac{1}{2}\pi$; it follows that $u_1, u_1 - u_2, u_1 - u_2 + u_3, \ldots$ are alternately greater and less than $\frac{1}{2}\pi$. Since $\frac{2\theta A}{2n+1}$ is arbitrarily small, for sufficiently great values of n, it thus appears that the differences

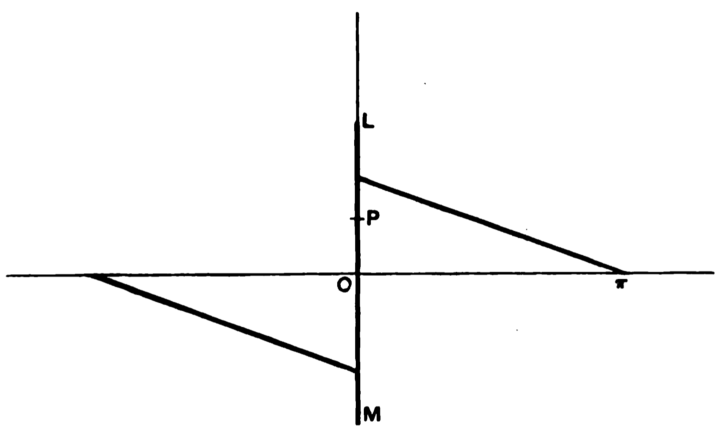
$$s_n\left(\frac{2\lambda\pi}{2n+1}\right) - s\left(\frac{2\lambda\pi}{2n+1}\right)$$

are alternately positive and negative for $\lambda = 1, 2, 3, \ldots$; and that for $\lambda = 1$, the difference is positive.



It thus appears that, for large values of n, the form of the curve $y = s_n(x)$ in the neighbourhood of the origin is as in the figure; consisting of a wave-form passing above and below the straight lines which represent $y = s(x)$. The first maximum on the right of the point $x=0$, has as its abscissa $x=\frac{2\pi}{2n+1}$, and its height above the point whose coordinates are $\frac{2\pi}{2n+1}$, $s\left(\frac{2\pi}{2n+1}\right)$ is nearly $\int_0^\pi \frac{\sin z}{z}\,dz - \frac{1}{2}\pi$, which is nearly independent of the value of n. The first minimum on the right of the point $x=0$, has for its abscissa $x=\frac{4\pi}{2n+1}$, and is at a depth approximately $\frac{1}{2}\pi - \int_0^{2\pi}\frac{\sin z}{z}\,dz$ below the corresponding point of the locus $y=s(x)$.

As n is continually increased, the abscissae of the maxima and minima of $s_n(x)-s(x)$ become indefinitely small, the magnitudes of these maxima and minima remaining however nearly unaltered. If a particular value of x be chosen, n can be so determined that $|s_n(x)-s(x)|$ is arbitrarily small, for such value of n, and for all greater values; but if a particular value of n be chosen, there is always a value of x, viz. $\frac{2\pi}{2n+1}$, for which $s_n(x)-s(x)$ is nearly equal to $\int_0^\pi \frac{\sin z}{z}\,dz - \frac{1}{2}\pi$.



The graphs $y=s_n(x)$, as n becomes indefinitely great, tend to the form given in the figure, which consists of the continuous curve formed by the straight lines of length $2\int_0^\pi \frac{\sin z}{z}\,dz\,(>\pi)$, through the points $x=0$, 2π, -2π, ..., and of the series of oblique straight lines which belong to the curve $y=s(x)$. The graph of the curve $y=s(x)=\lim\limits_{n=\infty} s_n(x)$ has been already given. The limit of the graphs of the curves $y=s_n(x)$, and the graph of the limit of $s_n(x)$ differ in the respect that, for the abscissae

$x=0,\ 2\pi,\ -2\pi, \ldots$, the former contains the continuous straight lines of length $2\int_0^\pi \frac{\sin z}{z}\,dz$, whereas the latter contains only the single points on the x-axis. Corresponding to any point P on the straight line LM through the origin, it is possible to determine an indefinite number of pairs of values of x and n, such that the distance of P from the point whose coordinates are x, $s_n(x)$, is less than an arbitrarily prescribed positive number ϵ. Thus the double limit $\lim_{n=\infty,\, x=0} s_n(x)$ is indeterminate between the limits of indeterminacy

$$\int_0^\pi \frac{\sin z}{z}\,dz,\quad -\int_0^\pi \frac{\sin z}{z}\,dz.$$

By letting n increase indefinitely, and x at the same time diminish to zero, in such a manner that nx has a as its limit, where a is any fixed positive number not exceeding π, we have as the particular value of $\lim_{n=\infty,\, x=0} s_n(x)$, or $\lim_{n=\infty} s_n\left(\frac{a}{n}\right)$, the number $\int_0^a \frac{\sin z}{z}\,dz$. It will be observed that the repeated limit $\lim_{x=0}\lim_{n=\infty} s_n(x)$ has the value $\frac{1}{2}\pi$, or $-\frac{1}{2}\pi$, according as x approaches its limit from the positive, or from the negative side. The repeated limit $\lim_{n=\infty}\lim_{x=0} s_n(x)$ has the value zero.

The distinction between the graph $y=s(x)$, which represents the series, and the limit to which the graphs $y=s_n(x)$ tend, is clear, if it be borne in mind that the limit $y=s(x)$ is obtained by the special mode of first fixing a value of x, and then letting n increase indefinitely; thus, for example, $s(0)=\lim_{n=\infty} s_n(0)=0$; whereas, as we have seen, $\lim_{n=\infty,\, x=0} s_n(x)$ is indeterminate between limits which have been found above. The difficulty which has been frequently felt in understanding how a series, of which the terms are continuous, such as the series here considered, can represent a function which is not continuous, will be removed if the point just explained be fully grasped*, that *the sum of the series at a point x is always taken to mean the limit obtained by first fixing the abscissa x, and then afterwards making the number of terms increase indefinitely.*

It has already been shewn in § 343, that the points $x=0,\ 2\pi,\ -2\pi, \ldots$, must be points of non-uniform continuity of the series; moreover, other examples have been already given, in which the peaks of the approximation curves $y=s_n(x)$ remain of finite height above the curve $y=s(x)$, however great n may be. That the portions of the limit of the graphs $y=s_n(x)$, in

* Some criticisms of Dirichlet's determination of the sum of a Fourier's series at a point of discontinuity, made by Schläfli, *Crelle's Journal*, vol. LXXII, and by Du Bois Reymond, *Math. Annalen*, vol. VII, where it is maintained that the sum of the series is indeterminate, are due to a lack of appreciation of this point.

the present case, have a length greater than π, the measure of discontinuity of the function, was pointed out by Willard Gibbs*.

The expression $\int_0^{(n+\frac{1}{2})x} \frac{\sin z}{z} dz - \frac{1}{2}\pi + \frac{2\theta A}{2n+1}$, which has been found above, for $s_n(x) - s(x)$, provided $0 < x \leqq b < 2\pi$, may be employed to shew that the series converges uniformly in any interval (a, b), such that $0 < a < b < 2\pi$. For, by choosing n so great that $\int_0^{(n+\frac{1}{2})x} \frac{\sin z}{z} dz$, for $x \geqq a$, differs from $\frac{1}{2}\pi$ by less than a prescribed number $\frac{1}{2}\epsilon$, which is possible on account of the convergence of the integral, and further choosing n so great that $\frac{2A}{2n+1} < \frac{1}{2}\epsilon$, it is seen that n can be chosen so great that, for the chosen value of n, and for all greater values, $|s_n(x) - s(x)| < \epsilon$, for all values of x in the interval (a, b). This expresses the fact that the series converges uniformly in the interval (a, b). It is clear that the smaller a is taken, the greater must be the value of n, so that $(n + \frac{1}{2})a$ may be sufficiently large to satisfy the requirement that $\left| \int_0^{(n+\frac{1}{2})a} \frac{\sin z}{z} dz - \frac{1}{2}\pi \right| < \frac{1}{2}\epsilon$; and that this value of n increases indefinitely as a is indefinitely diminished. This is a verification of the fact that the convergence of the series is non-uniform at the point $x = 0$.

442. Let $f(x)$ be defined for the interval $(0, \pi)$, by the specifications

$$f(x) = c, \text{ for } 0 \leqq x < \tfrac{1}{2}\pi,$$

and

$$f(x) = -c, \text{ for } \tfrac{1}{2}\pi \leqq x \leqq \pi.$$

To find the sine series for this function, we have

$$\int_0^\pi f(x) \sin nx\,dx = c\int_0^{\frac{1}{2}\pi} \sin nx\,dx - c\int_{\frac{1}{2}\pi}^\pi \sin nx\,dx$$

$$= \frac{c}{n}(\cos n\pi - 2\cos \tfrac{1}{2}n\pi + 1).$$

This integral vanishes if n is odd, and also if n is a multiple of 4, but i $n = 4m + 2$, it has the value $4c/n$. The series is therefore

$$\frac{8c}{\pi}(\tfrac{1}{2}\sin 2x + \tfrac{1}{6}\sin 6x + \tfrac{1}{10}\sin 10x + \ldots).$$

For unrestricted values of x, this series represents the ordinates of the series of straight lines in the next figure, except that it vanishes at the points

$$0,\ \tfrac{1}{2}\pi,\ \pi,\ -\tfrac{1}{2}\pi,\ -\pi,\ \ldots.$$

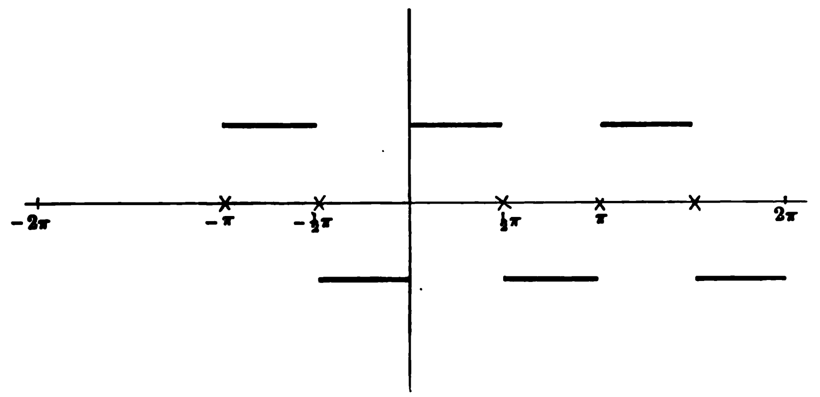
It will be observed that, if the meaning of $f(x)$ be extended, so that it denotes

* See an interesting discussion on this subject in *Nature*, vol. LVIII, 1898, pp. 544, 569, vol. LIX, 1899, pp. 200, 271, 319, 606, vol. LX, pp. 52, 100, in which Michelson, Love, Gibbs, Baker and Poincaré took part.

the sum of the sine series for every value of x for which that sum is continuous, then at the point π, for example,

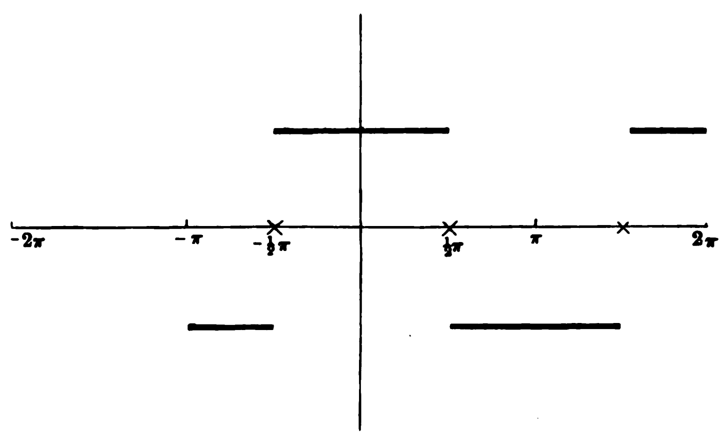
$$f(\pi+0)=c,\ f(\pi-0)=-c,$$

and the series represents at the point π, the arithmetic mean of these two values.



In a similar manner, we find that the function defined for the interval $(0, \pi)$ as before, is represented, for the interval $(0, \pi)$, by the cosine series

$$\frac{4c}{\pi}(\cos x - \tfrac{1}{3}\cos 3x + \tfrac{1}{5}\cos 5x - \ldots).$$



For unrestricted values of x, the series represents the ordinates of the straight lines in the figure, except that its sum vanishes at the points

$$\tfrac{1}{2}\pi,\ -\tfrac{1}{2}\pi,\ \tfrac{3}{2}\pi,\ \ldots.$$

443. Let $f(x)=x$, for $0 \leqq x \leqq \tfrac{1}{2}\pi$,

and $f(x)=\pi-x$, for $\tfrac{1}{2}\pi \leqq x \leqq \pi$.

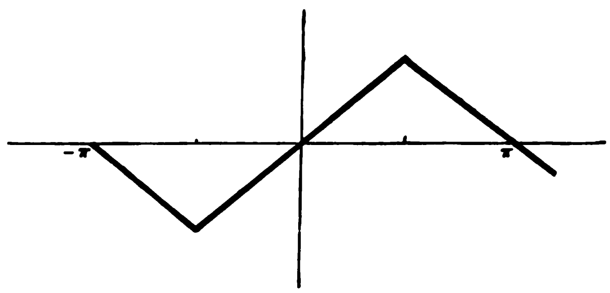
In this case we find that

$$\int_0^{\pi} f(x)\sin nx\,dx = \int_0^{\frac{1}{2}\pi} x\sin nx\,dx + \int_{\frac{1}{2}\pi}^{\pi} (\pi - x)\sin nx\,dx$$

$$= \frac{2}{n^2}\sin \tfrac{1}{2}n\pi.$$

Hence the sine series is

$$\frac{4}{\pi}\left(\sin x - \frac{1}{3^2}\sin 3x + \frac{1}{5^2}\sin 5x - \dots\right).$$

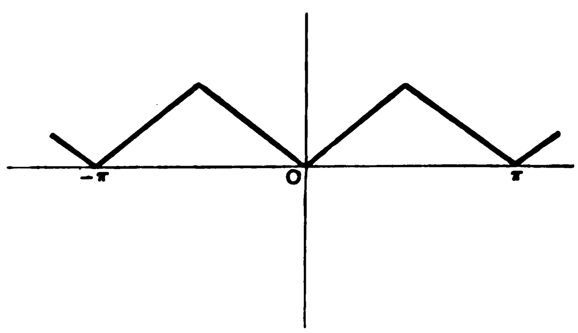


For general values of x, the series represents the ordinates of the line in the figure. The broken line in the interval $(-\pi, \pi)$ is repeated indefinitely in both directions.

The cosine series which represents the same function for the interval $(0, \pi)$, will be found to be

$$\frac{1}{4}\pi - \frac{2}{\pi}\left(\cos 2x + \frac{1}{3^2}\cos 6x + \frac{1}{5^2}\cos 10x + \dots\right).$$

This series represents for general values of x, the ordinates of the line in the following figure. As before, the broken line in the interval $(-\pi, \pi)$ is to be repeated indefinitely in both directions.



EXAMPLES.

1. Prove that the series

$$\sin x - \tfrac{1}{2}\sin 2x + \tfrac{1}{3}\sin 3x - \tfrac{1}{4}\sin 4x + \ldots$$

represents, for the interior of the interval $(-\pi, \pi)$, the function $\tfrac{1}{2}x$.

For any value of x which is not a multiple of π, the series represents $\tfrac{1}{2}(x - 2k\pi)$, where k is a positive or negative integer so chosen that $x - 2k\pi$ lies between π and $-\pi$. The sum of the series vanishes for all values of x which are multiples of π.

2. Prove that the series

$$\cos x - \tfrac{1}{4}\cos 2x + \tfrac{1}{9}\cos 3x - \tfrac{1}{16}\cos 4x + \ldots$$

represents the function $\tfrac{1}{12}\pi^2 - \tfrac{1}{4}x^2$, for the interval $(-\pi, \pi)$.

3. Prove that

$$\tfrac{1}{4}\pi = \sin x + \tfrac{1}{3}\sin 3x + \tfrac{1}{5}\sin 5x + \ldots, \qquad \text{for } 0 < x < \pi;$$

$$\tfrac{1}{4}\pi = \cos x - \tfrac{1}{3}\cos 3x + \tfrac{1}{5}\cos 5x - \ldots, \qquad \text{for } -\tfrac{1}{2}\pi < x < \tfrac{1}{2}\pi.$$

4. Prove that

$$\frac{1}{4}\pi x = \sin x - \frac{1}{3^2}\sin 3x + \frac{1}{5^2}\sin 5x - \ldots, \qquad \text{for } -\frac{1}{2}\pi \leqq x \leqq \frac{1}{2}\pi.$$

5. Prove that

$$e^{kx} = \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{n}{k^2+n^2}(1 - e^{k\pi}\cos n\pi)\sin nx, \qquad \text{for } 0 < x < \pi,$$

$$e^{kx} = \frac{e^{k\pi}-1}{k\pi} + \frac{2k}{\pi}\sum_{n=1}^{\infty}\frac{e^{k\pi}\cos n\pi - 1}{k^2+n^2}\cos nx, \qquad \text{for } 0 \leqq x \leqq \pi.$$

6. Prove that

$$\frac{\pi}{2}\cdot\frac{\sin kx}{\sin k\pi} = \frac{\sin x}{1^2-k^2} - \frac{2\sin 2x}{2^2-k^2} + \frac{3\sin 3x}{3^2-k^2} - \ldots, \qquad \text{where } 0 \leqq x < \pi,$$

$$\frac{\pi}{2}\cdot\frac{\cos kx}{\sin k\pi} = \frac{1}{2k} - \frac{k\cos x}{k^2-1^2} + \frac{k\cos 2x}{k^2-2^2} - \ldots, \qquad \text{where } 0 \leqq x \leqq \pi;$$

k not being integral.

7. Prove that

$$\frac{\pi}{2}\frac{\sinh kx}{\sinh k\pi} = \frac{\sin x}{1^2+k^2} - \frac{2\sin 2x}{2^2+k^2} + \frac{3\sin 3x}{3^2+k^2} - \ldots, \qquad \text{where } 0 \leqq x < \pi;$$

$$\frac{\pi}{2k}\frac{\cosh k(\pi - x)}{\sinh k\pi} = \frac{1}{2k^2} + \frac{\cos x}{1^2+k^2} + \frac{\cos 2x}{2^2+k^2} + \frac{\cos 3x}{3^2+k^2} + \ldots, \qquad \text{where } 0 \leqq x \leqq \pi.$$

DIRICHLET'S INTEGRAL.

444. It has been shewn in § 440, that the sum S_{2n+1}, of the first $2n+1$ terms of Fourier's series, is of the form

$$S_{2n+1} = \frac{1}{\pi}\int_{-\frac{1}{2}(\pi+x)}^{\frac{1}{2}(\pi-x)} f(x+2z)\frac{\sin mz}{\sin z}\,dz,$$

where $\qquad m = 2n+1.$

The function $f(x)$ has hitherto been defined for the interval $(-\pi, \pi)$ only; we may now, as a matter of convenience, extend the definition of $f(x)$ to values of x which do not lie in this interval.

We shall assume that, for all values of x which are not multiples of π, the functional values are defined so that, for all such values of x, the condition $f(x+2\pi)=f(x)$ is satisfied. In case $f(\pi)=f(-\pi)$, we may, if we please, suppose that $f(\pi)=f(\pm k\pi)$ for all integral values of k; but in any case it is indifferent what values are assigned to the function at the points $\pm k\pi$. Whenever $f(\pi)$ and $f(-\pi)$ are unequal, one at least of the points $\pi, -\pi$ is certainly a point of discontinuity of $f(x)$,

for $$f(\pi+0)=f(-\pi+0), \text{ and } f(\pi-0)=f(-\pi-0),$$

in accordance with our extended definition of $f(x)$; it is accordingly impossible that both of the sets of conditions

$$f(\pi)=f(\pi+0)=f(\pi-0),$$
$$f(-\pi)=f(-\pi+0)=f(-\pi-0)$$

can be satisfied, if $f(\pi)$ and $f(-\pi)$ be unequal.

Let us now write $f(x+2z)=F(z)$, then the function $F(z)$ is periodic in z, with period π, except possibly for those values of z for which $x+2z$ is a multiple of π; this exception is, however, immaterial, since the values of integrals are unaffected by alteration of the functional values at single points.

We may now write S_m in the form

$$S_m=\frac{1}{\pi}\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} F(z)\frac{\sin mz}{\sin z}\,dz,$$

which may also be written in the form

$$S_m=\frac{1}{\pi}\int_0^{\frac{1}{2}\pi} F(z)\frac{\sin mz}{\sin z}\,dz+\frac{1}{\pi}\int_0^{\frac{1}{2}\pi} F(-z)\frac{\sin mz}{\sin z}\,dz.$$

It thus appears that the investigation of the limiting value of S_m turns upon the existence and value of the limit, when m is indefinitely increased, of an integral of the form

$$\int_0^{\frac{1}{2}\pi} F(z)\frac{\sin mz}{\sin z}\,dz;$$

the second integral in the expression for S_m being essentially of the same form.

This integral is known as *Dirichlet's integral*, the term being, however, generally applied to the more general form

$$\int_0^{a} F(z)\frac{\sin mz}{\sin z}\,dz,$$

where a is such that $$0<a\leqq\tfrac{1}{2}\pi.$$

The term Dirichlet's integral is also frequently applied to the closely allied expression

$$\int_0^a F(z)\frac{\sin z}{z}\,dz, \text{ where } 0 < a \leqq \tfrac{1}{2}\pi.$$

It will be shewn that, with certain assumptions as to the nature of the function $f(x)$, the integrals

$$\int_0^{\frac{1}{2}\pi} F(z)\frac{\sin mz}{\sin z}\,dz,$$

$$\int_0^{\frac{1}{2}\pi} F(-z)\frac{\sin mz}{\sin z}\,dz,$$

converge to the values

$$\frac{\pi}{2}F(+0),\ \ \frac{\pi}{2}F(-0),$$

respectively, as m is increased indefinitely; so that S_m converges to the value

$$\tfrac{1}{2}\{f(x+0)+f(x-0)\}.$$

At a point of continuity of $f(x)$, this agrees with $f(x)$. When $x = \pm\pi$, S_m converges, under certain assumptions, to the value

$$\tfrac{1}{2}\{f(\pi-0)+f(-\pi+0)\}.$$

DIRICHLET'S INVESTIGATION OF FOURIER'S SERIES.

445. As a preliminary to the consideration of Dirichlet's integral, some properties of the integral

$$\int_0^{\frac{\pi}{2}} \frac{\sin mz}{\sin z}\,dz$$

are required.

We have

$$\int_0^{\frac{\pi}{2}} \frac{\sin mz}{\sin z}\,dz = \int_0^{\frac{\pi}{2}} [1 + 2\cos 2z + 2\cos 4z + \dots + 2\cos 2nz]\,dz = \frac{\pi}{2}.$$

If we divide the interval $\left(0, \frac{\pi}{2}\right)$, of integration, into the portions

$$\left(0, \frac{\pi}{m}\right),\ \left(\frac{\pi}{m}, \frac{2\pi}{m}\right),\ \dots\ \left(\frac{r\pi}{m}, \frac{\overline{r+1}\,\pi}{m}\right),\ \dots\ \left(\frac{n\pi}{m}, \frac{\pi}{2}\right),$$

we see that, in these portions, the integrand $\frac{\sin mz}{\sin z}$ has alternately positive and negative signs; thus if we write

$$\rho_{r-1} = (-1)^{r-1} \int_{\frac{\overline{r-1}\pi}{m}}^{\frac{r\pi}{m}} \frac{\sin mz}{\sin z} dz,$$

$$\rho_n = (-1)^n \int_{\frac{n\pi}{m}}^{\frac{\pi}{2}} \frac{\sin mz}{\sin z} dz,$$

we have $$\frac{\pi}{2} = \rho_0 - \rho_1 + \rho_2 + \ldots + (-1)^{r-1} \rho_{r-1} + \ldots + (-1)^n \rho_n,$$

where all the ρ's are positive.

In ρ_{r-1}, $\sin mz$ is always of the same sign, and $\frac{1}{\sin z}$ is monotone and decreases as z increases, hence

$$\rho_{r-1} < (-1)^{r-1} \frac{1}{\sin \frac{\overline{r-1}\pi}{m}} \int_{\frac{\overline{r-1}\pi}{m}}^{\frac{r\pi}{m}} \sin mz \,.\, dz < \frac{2}{m} \operatorname{cosec} \frac{\overline{r-1}\pi}{m};$$

and similarly $$\rho_{r-1} > \frac{2}{m} \operatorname{cosec} \frac{r\pi}{m}.$$

It follows that $$\rho_{r-1} > \frac{2}{m} \operatorname{cosec} \frac{r\pi}{m} > \rho_r.$$

For ρ_n, we have

$$\frac{1}{m} \operatorname{cosec} \frac{n\pi}{m} > \rho_n > \frac{1}{m},$$

hence $$\rho_{n-1} > \frac{2}{m} \operatorname{cosec} \frac{n\pi}{m} > \rho_n.$$

It follows that, if $2p < n$,

$$\frac{\pi}{2} < \rho_0 - \rho_1 + \rho_2 - \ldots + \rho_{2p},$$

and $$\frac{\pi}{2} > \rho_0 - \rho_1 + \rho_2 - \ldots - \rho_{2p-1}.$$

Let us suppose that the function $F(z)$ has a finite upper limit, for the values of z such that $0 \leqq z \leqq \frac{1}{2}\pi$, and further, that it is in the whole interval positive and monotone, and such that it never increases as z increases; it is consequently an integrable function.

In the integral

$$\int_0^\alpha F(z) \frac{\sin mz}{\sin z} dz,$$

where $\alpha \leqq \frac{1}{2}\pi$, we proceed to divide the interval of integration as in the case of

$$\int_0^{\frac{\pi}{2}} \frac{\sin mz}{\sin z} dz$$

into alternately positive and negative portions; thus if

$$s_{r-1} = (-1)^{r-1} \int_{\frac{\overline{r-1}\pi}{m}}^{\frac{r\pi}{m}} F(z) \frac{\sin mz}{\sin z} dz,$$

$$s_q = (-1)^q \int_{\frac{q\pi}{m}}^{a} F(z) \frac{\sin mz}{\sin z} dz,$$

where q is a positive integer such that

$$\frac{q\pi}{m} < a \leqq \frac{\overline{q+1}\pi}{m},$$

we have

$$\int_0^a F(z) \frac{\sin mz}{\sin z} dz = s_0 - s_1 + s_2 - \dots + (-1)^{r-1} s_{r-1} + \dots + (-1)^q s_q,$$

where $s_0, s_1, s_2, \dots s_q$ are all positive. On account of the supposition made as regards $F(z)$, we have

$$\rho_{r-1} F\left(\frac{\overline{r-1}\pi}{m}\right) \geqq s_{r-1} \geqq \rho_{r-1} F\left(\frac{r\pi}{m}\right), \text{ and } s_q \leqq \rho_q F\left(\frac{q\pi}{m}\right).$$

From these inequalities it follows that

$$s_{r-1} \geqq \rho_{r-1} F\left(\frac{r\pi}{m}\right) > \rho_r F\left(\frac{r\pi}{m}\right) > s_r;$$

and this holds for all values of r from 1 to q.

We have consequently the result, that

$$U \equiv \int_0^a F(z) \frac{\sin mz}{\sin z} dz,$$

is less than $s_0 - s_1 + s_2 - \dots - s_{2p-1} + s_{2p}$,

and greater than $s_0 - s_1 + s_2 - \dots - s_{2p-1}$, where $2p \leqq q$.

From these inequalities, with the help of those obtained above, we have

$$U > (\rho_0 - \rho_1) F\left(\frac{\pi}{m}\right) + (\rho_2 - \rho_3) F\left(\frac{3\pi}{m}\right) + \dots + (\rho_{2p-2} - \rho_{2p-1}) F\left(\frac{\overline{2p-1}\pi}{m}\right)$$

$$> F\left(\frac{2p\pi}{m}\right)(\rho_0 - \rho_1 + \rho_2 - \rho_3 + \dots + \rho_{2p-2} - \rho_{2p-1});$$

also $$U < \rho_0 F(+0) - F\left(\frac{2p\pi}{m}\right)(\rho_1 - \rho_2 + \rho_3 - \dots - \rho_{2p}).$$

On using the theorems which have been proved relating to the ρ's, we obtain

$$U > F\left(\frac{2p\pi}{m}\right)\left(\frac{\pi}{2} - \rho_{2p}\right)$$

and $$U < \rho_0 \left\{F(+0) - F\left(\frac{2p\pi}{m}\right)\right\} + \left(\frac{\pi}{2} + \rho_{2p}\right) F\left(\frac{2p\pi}{m}\right);$$

where, in accordance with the supposition made, p is any integer such that

$$2p \leqq q < \frac{m\alpha}{\pi}.$$

Now let m and p both increase indefinitely, but in such a way that $\frac{2p}{m}$ has the limit zero. Since

$$\rho_{2p} < \frac{2}{m \sin \frac{2p\pi}{m}} < \frac{1}{p\pi} \cdot \frac{\frac{2p\pi}{m}}{\sin \frac{2p\pi}{m}},$$

we see that ρ_{2p} has zero for its limit; and hence

$$F\left(\frac{2p\pi}{m}\right)\left(\frac{\pi}{2} - \rho_{2p}\right)$$

has $\frac{\pi}{2} F(+0)$ for its limit. Again

$$\rho_0 < \frac{\pi}{2} + \rho_1 < \frac{\pi}{2} + \frac{2}{\pi} \frac{\frac{\pi}{m}}{\sin \frac{\pi}{m}};$$

and hence ρ_0 has a limiting value not greater than $\frac{\pi}{2} + \frac{2}{\pi}$. It follows that

$$\rho_0 \left\{F(+0) - F\left(\frac{2p\pi}{m}\right)\right\} + \left(\frac{\pi}{2} + \rho_{2p}\right) F\left(\frac{2p\pi}{m}\right)$$

has for its limit the value $\frac{\pi}{2} F(+0)$.

It has been proved that U lies between two numbers, each of which has $\frac{\pi}{2} F(+0)$ for limit, when m and p are indefinitely increased in such a way that $\frac{2p}{m}$ has the limit zero; hence the limit of

$$U \equiv \int_0^\alpha F(z) \frac{\sin mz}{\sin z} dz$$

is

$$\frac{\pi}{2} F(+0),$$

where α is such that

$$0 < \alpha \leqq \tfrac{1}{2}\pi.$$

It follows, as a corollary from this theorem, that

$$\int_\beta^\alpha F(z) \frac{\sin mz}{\sin z} dz$$

has the limit zero, when m is indefinitely increased; where α, β are two fixed numbers, such that $0 < \beta < \alpha \leqq \frac{1}{2}\pi$.

446. We have now seen that, if $F(z)$ be a limited and positive function which never increases as z increases from 0 to $\frac{1}{2}\pi$, the integral

$$\int_0^{\frac{\pi}{2}} F(z)\frac{\sin mz}{\sin z}\,dz$$

converges to the value $\frac{\pi}{2}F(+0)$, as m is increased indefinitely. The function $F(z)$ may be freed from the condition that it must be positive in the whole interval. For if $F\left(\frac{\pi}{2}\right)$ is negative, we may apply the theorem to the function $C+F(z)$, where the constant C is chosen so that

$$C+F\left(\frac{\pi}{2}\right)$$

is positive; thus

$$\int_0^{\frac{\pi}{2}} \{C+F(z)\}\frac{\sin mz}{\sin z}\,dz$$

converges to the limit $\frac{\pi}{2}\{C+F(+0)\}$.

Now

$$C\int_0^{\frac{\pi}{2}} \frac{\sin mz}{\sin z}\,dz$$

converges to the limit $\frac{\pi}{2}C$; hence $\int_0^{\frac{\pi}{2}} F(z)\frac{\sin mz}{\sin z}\,dz$ converges to $\frac{\pi}{2}F(+0)$, where $F(z)$ is not restricted to be positive.

Again, the theorem holds for a function $F(z)$ which is monotone and never diminishes; for we can apply the theorem to the monotone function $-F(z)$ which never increases.

The theorem has now been established, that *if $F(z)$ be any limited, monotone function, defined for the interval* $(0, \frac{1}{2}\pi)$, *then*

$$\int_0^{\frac{\pi}{2}} F(z)\frac{\sin mz}{\sin z}\,dz$$

converges, as the odd integer m is increased indefinitely, to the value $\frac{\pi}{2}F(+0)$.

The theorem also holds if the upper limit of the integral be any fixed number α, such that $0<\alpha\leqq\frac{1}{2}\pi$.

It has been shewn in § 195, that any function with limited total fluctuation is expressible as the difference of two monotone functions. Hence the results which have been established can be immediately extended to the case of functions of this class. We have, therefore, the theorem that, *if*

$F(z)$ be a function defined for the interval $(0, \frac{1}{2}\pi)$, and with limited total fluctuation, then the integrals

$$\int_0^a F(z)\frac{\sin mz}{\sin z}dz, \quad \int_a^\beta F(z)\frac{\sin mz}{\sin z}dz,$$

where $$0 < a \leqq \tfrac{1}{2}\pi, \quad 0 < a < \beta \leqq \tfrac{1}{2}\pi,$$

converge, as the odd integer m is increased indefinitely, to the values $\frac{\pi}{2}F(+0)$, 0 respectively.

If we apply this result to the two integrals contained in the expression for S_m, the sum of the first $2n+1$ terms in Fourier's series, we obtain the theorem that, *if $f(x)$ be a function with limited total fluctuation, defined for the interval $(-\pi, \pi)$, the sum of $2n+1$ terms of the series*

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x')\,dx' + \sum_{n=1}\left\{\frac{1}{\pi}\cos nx\int_{-\pi}^{\pi} f(x')\cos nx'\,dx' + \frac{1}{\pi}\sin nx\int_{-\pi}^{\pi} f(x')\sin nx'\,dx'\right\}$$

converges, as n is indefinitely increased, to the value

$$\tfrac{1}{2}\{f(x+0)+f(x-0)\}.$$

It will be remembered that a function with limited total fluctuation is essentially integrable, in accordance with Riemann's definition; and that it can have discontinuities of the first kind only, so that at every point the functional limits $f(x+0)$, $f(x-0)$ exist.

In the case $x = \pm\pi$, the limit to which the sum of the series converges is

$$\tfrac{1}{2}\{f(\pi-0)+f(-\pi+0)\}.$$

At a point x of continuity of the function $f(x)$, the limiting sum of the series is $f(x)$; at a point of discontinuity of $f(x)$, the limiting sum of the series agrees with the value of the function at the point only if

$$f(x) = \tfrac{1}{2}\{f(x+0)+f(x-0)\}.$$

At the points π, $-\pi$, the limiting sum of the series agrees with the value of the function only if $f(\pi)$, or $f(-\pi)$, is equal to

$$\tfrac{1}{2}\{f(\pi-0)+f(-\pi+0)\}.$$

447. It is now clear in what sense the given function $f(x)$ is represented by the corresponding Fourier's series. The representation is necessarily complete for all points at which the function is continuous, with the possible exception of the end-points $\pm\pi$, which cannot both be points of continuity of the extended function, unless $f(\pi) = f(-\pi)$. At a point of discontinuity, or at an end-point $\pm\pi$, the series represents the function only if the functional value is properly chosen in relation to the functional limits at the point; in the case of the end-points these functional limits are those of the periodic function obtained by extension of the given function beyond the domain for which it was at first defined, this extension being such that $f(x)=f(x+2\pi)$, as explained in § 444.

The functions with limited total fluctuation include, as a particular case, functions which satisfy the following conditions:—

(1) The function is continuous in its domain at every point, with the exception of a finite number of points at which it may have ordinary discontinuities, (2) the domain may be divided into a finite number of parts, such that in any one of them the function is monotone; or in accordance with the more usual but less exact expression, the function has only a finite number of maxima and minima in its domain.

These conditions are known as *Dirichlet's conditions*, and his proof, in its original form, applied to the case only of functions which satisfy these conditions.

448. Dirichlet extended his results to the case in which there are a finite number of points in the domain $(-\pi, \pi)$ in the neighbourhood of which $|f(x)|$ has no upper limit. In this case the Fourier's series must be so interpreted that the integrals in the coefficients are the improper integrals

$$\int_{-\pi}^{\pi} f(x)\,dx, \quad \int_{-\pi}^{\pi} {\cos \atop \sin} nx\, f(x)\,dx,$$

the function being such that these improper integrals exist. From our somewhat more general point of view, we shall suppose that the function $f(x)$ is such that, when arbitrarily small neighbourhoods of these infinite singularities are excluded from the interval $(-\pi, \pi)$, in the remaining part of the interval $f(x)$ is of limited total fluctuation; and further it will be assumed that the improper integral

$$\int_{-\pi}^{\pi} f(x)\,dx$$

exists, and is absolutely convergent. Under these conditions, it can be shewn that the theorems still hold, that the integrals

$$\int_{\alpha}^{\beta} F(z)\frac{\sin mz}{\sin z}\,dz, \text{ for } 0<\alpha<\beta \leqq \tfrac{1}{2}\pi,$$

and

$$\int_{0}^{\alpha} F(z)\frac{\sin mz}{\sin z}\,dz, \text{ for } 0<\alpha \leqq \tfrac{1}{2}\pi,$$

converge to zero, and to $\frac{\pi}{2}F(+0)$, respectively, as m is increased indefinitely.

If, between α and β, there is a point c in whose neighbourhood $|F(z)|$ has no upper limit,

$$\int_{\alpha}^{\beta} F(z)\frac{\sin mz}{\sin z}\,dz$$

is interpreted as the limit of

$$\int_a^{c-\delta} F(z) \frac{\sin mz}{\sin z} dz + \int_{c+\epsilon}^{\beta} F(z) \frac{\sin mz}{\sin z} dz,$$

where δ, ϵ have, independently of one another, the limit zero; assuming that such limit exists.

Let $\delta' < \delta$, then

$$\left| \left[\int_a^{c-\delta'} - \int_a^{c-\delta} \right] F(z) \frac{\sin mz}{\sin z} dz \right| < \operatorname{cosec} \alpha \int_{c-\delta}^{c-\delta'} | F(z) | dz;$$

where the expression on the right-hand side is arbitrarily small, on account of the absolute convergence of the integral of $F(z)$, and is independent of the value of m.

Now, if $\int F(z)\, dz$ converges absolutely at the point c, we can choose δ so small that, for every $\delta' < \delta$,

$$\operatorname{cosec} \alpha \int_{c-\delta}^{c-\delta'} | F(z) | dz$$

is arbitrarily small; hence the integral

$$\int_a^{c-\delta} F(z) \frac{\sin mz}{\sin z} dz,$$

for a fixed m, converges to a definite value, as δ converges to zero. Similarly it can be shewn that

$$\int_{c+\epsilon}^{\beta} F(z) \frac{\sin mz}{\sin z} dz$$

converges to a definite value, as ϵ converges to zero. It has thus been shewn that

$$\int_a^{\beta} F(z) \frac{\sin mz}{\sin z} dz = \lim_{\delta=0} \int_a^{c-\delta} F(z) \frac{\sin mz}{\sin z} dz + \lim_{\epsilon=0} \int_{c+\epsilon}^{\beta} F(z) \frac{\sin mz}{\sin z} dz = \psi_1(m) + \psi_2(m);$$

and we have now to shew that $\psi_1(m)$, $\psi_2(m)$ converge to zero as m is increased indefinitely. It has been already seen that δ may be so chosen that, for all values of m,

$$\left| \int_a^{c-\delta} F(z) \frac{\sin mz}{\sin z} dz - \psi_1(m) \right| < \eta,$$

where η is a fixed arbitrarily small positive number. Now, for a fixed value of δ, m_1 may be chosen so great that, if $m \geqq m_1$,

$$\left| \int_a^{c-\delta} F(z) \frac{\sin mz}{\sin z} dz \right| < \zeta,$$

where ζ is arbitrarily small; hence, if $m \geqq m_1$,

$$|\psi_1(m)| < \eta + \zeta,$$

and therefore $\psi_1(m)$ converges to the limit zero; similarly $\psi_2(m)$ converges to the limit zero.

If, between α and β, there are any finite number of points such as c, we may divide the domain (α, β) into a finite number of parts, such that each part contains only one such point as c, and apply the above result to each of the integrals which are taken through one such part.

The integral $\int_0^a F(z)\frac{\sin mz}{\sin z}\,dz$ can be divided into two parts

$$\int_0^{\alpha_1} F(z)\frac{\sin mz}{\sin z}\,dz + \int_{\alpha_1}^a F(z)\frac{\sin mz}{\sin z}\,dz,$$

where α_1 is so chosen that all the points of infinite discontinuity of $F(z)$ are in (α_1, α); we thus see that $\int_0^a F(z)\frac{\sin mz}{\sin z}\,dz$ converges to $\frac{\pi}{2}F(+0)$, when m is indefinitely increased.

It has now been shewn that:—*if $f(x)$ be such that, when the arbitrarily small neighbourhoods of a finite number of points at which $|f(x)|$ has no upper limit have been excluded, $f(x)$ becomes a function with limited total fluctuation, then the Fourier's series*

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x')\,dx' + \Sigma\frac{1}{\pi}\int_{-\pi}^{\pi} f(x')\cos n(x - x')\,dx'$$

converges to the value $\frac{1}{2}\{f(x+0) + f(x-0)\}$, at every point in $(-\pi, \pi)$, except at the points of infinite discontinuity of the function, provided the improper integral $\int_{-\pi}^{\pi} f(x)\,dx$ exists, and is absolutely convergent.

APPLICATION OF THE SECOND MEAN VALUE THEOREM.

449. An alternative method of investigation of the limit to which the sum of Fourier's series, corresponding to a function with limited total fluctuation, converges is obtained by employing the second mean value theorem*. As before, we employ the known result that any function with limited total fluctuation is expressible as the difference of two functions each of which is monotone and does not decrease as the variable increases; and it is therefore sufficient to consider the case of such a monotone function.

* This method was first employed by Bonnet, who used his form of the mean value theorem, see the *Mémoires des Savants étrangers* of the Belgian Academy, vol. XXIII. The method is also used by C. Neumann, see his work *Ueber die nach Kreis- Kugel- und Cylinderfunctionen fortschreitenden Reihen*; also by Jordan, *Cours d'Analyse*, vol. II, where it is applied to the case of functions with limited total fluctuation. The method is employed, and discussed in great detail, in Dini's work *Sopra le Serie di Fourier*.

(1) We have, as before,

$$\int_0^{\frac{\pi}{2}} \frac{\sin mz}{\sin z}\, dz = \frac{\pi}{2}.$$

(2) If
$$0 < \alpha < \beta \leqq \frac{\pi}{2},$$

$$\int_\alpha^\beta \frac{\sin mz}{\sin z}\, dz = \frac{1}{\sin\alpha}\int_\alpha^\gamma \sin mz\, dz + \frac{1}{\sin\beta}\int_\gamma^\beta \sin mz\, dz,$$

therefore

$$\left|\int_\alpha^\beta \frac{\sin mz}{\sin z}\, dz\right| < \frac{2}{m}(\operatorname{cosec}\alpha + \operatorname{cosec}\beta) < \frac{4}{m}\operatorname{cosec}\alpha;$$

here, in accordance with the second mean value theorem, γ is some number which lies between α and β.

It follows that $\int_\alpha^\beta \frac{\sin mz}{\sin z}\, dz$ converges to zero, as m is increased indefinitely.

In a precisely similar manner, it appears that $\int_\alpha^\beta \frac{\sin mz}{z}\, dz$ is, in absolute value, $< \frac{4}{m\alpha}$, and thus the value of the integral converges to zero, as m is indefinitely increased.

(3) If $\alpha \geqq 0$,
$$\left|\int_\alpha^\infty \frac{\sin\theta}{\theta}\, d\theta\right| \leqq \frac{\pi}{2}.$$

For
$$\left|\int_\alpha^h \frac{\sin\theta}{\theta}\, d\theta\right| < \frac{2}{\alpha} + \frac{2}{h},$$

and therefore
$$\left|\lim_{h=\infty}\int_\alpha^h \frac{\sin\theta}{\theta}\, d\theta\right| \leqq \frac{2}{\alpha};$$

hence, if $\alpha \geqq \pi$,
$$\left|\int_\alpha^\infty \frac{\sin\theta}{\theta}\, d\theta\right| \leqq \frac{2}{\pi} < \frac{\pi}{2}.$$

Now
$$\frac{d}{d\alpha}\int_\alpha^\infty \frac{\sin\theta}{\theta}\, d\theta = -\frac{\sin\alpha}{\alpha}, \text{ when } \alpha > 0;$$

hence $\int_\alpha^\infty \frac{\sin\theta}{\theta}\, d\theta$ diminishes as α increases, provided $0 < \alpha < \pi$; therefore since

$$\int_0^\infty \frac{\sin\theta}{\theta}\, d\theta = \frac{\pi}{2},$$

we have
$$\left|\int_\alpha^\infty \frac{\sin\theta}{\theta}\, d\theta\right| \leqq \frac{\pi}{2},$$

if $\alpha < \pi$, and $< \frac{2}{\pi}$, if $\alpha \geqq \pi$.

From this result, it follows that

$$\left|\int_\alpha^\beta \frac{\sin\theta}{\theta}\,d\theta\right| \leqq \pi, \qquad \text{where } 0 \leqq \alpha < \beta.$$

450. After having established these preliminary theorems, we proceed to consider

$$\int_0^{\frac{\pi}{2}} F(z)\frac{\sin mz}{\sin z}\,dz,$$

where $F(z)$ is monotone, and does not diminish as z increases. We have

$$\int_0^{\frac{\pi}{2}} F(z)\frac{\sin mz}{\sin z}\,dz = F(+0)\int_0^{\frac{\pi}{2}}\frac{\sin mz}{\sin z}\,dz + \int_0^\mu \{F(z)-F(+0)\}\frac{\sin mz}{\sin z}\,dz$$
$$+\int_\mu^{\frac{\pi}{2}}\{F(z)-F(+0)\}\frac{\sin mz}{\sin z}\,dz,$$

where μ is fixed. On applying the second mean value theorem, we have

$$\int_0^{\frac{\pi}{2}} F(z)\frac{\sin mz}{\sin z}\,dz - \frac{\pi}{2}F(+0)$$
$$=\int_0^\mu G(z)\frac{\sin mz}{z}\,dz + \{F(\mu+0)-F(+0)\}\int_\mu^{\xi_1}\frac{\sin mz}{\sin z}\,dz$$
$$+\left\{F\left(\frac{\pi}{2}-0\right)-F(+0)\right\}\int_{\xi_1}^{\frac{\pi}{2}}\frac{\sin mz}{\sin z}\,dz,$$

where ξ_1 is between μ and $\frac{\pi}{2}$, and $G(z)=\{F(z)-F(+0)\}\frac{z}{\sin z}$. Also $G(z)$ increases as z increases from 0 to μ.

Again $$\int_0^\mu G(z)\frac{\sin mz}{z}\,dz = G(\mu)\int_\xi^\mu \frac{\sin mz}{z}\,dz = G(\mu)\int_{m\xi}^{m\mu}\frac{\sin\theta}{\theta}\,d\theta,$$

where $0 \leqq \xi < \mu$.

Although it is unnecessary for the purpose of the present investigation, it may be remarked that we cannot have $\xi = 0$, which would involve the equality

$$\int_0^\mu \{G(\mu)-G(z)\}\frac{\sin mz}{z}\,dz = 0.$$

This is impossible unless $G(z)=G(\mu)$, for all values of z in the interval $(0, \mu)$. For let $z=\frac{z'}{m}$, then

$$\int_0^\mu \{G(\mu)-G(z)\}\frac{\sin mz}{z}\,dz$$
$$=\int_0^\pi \left\{G(\mu)-G\left(\frac{z'}{m}\right)\right\}\frac{\sin z'}{z'}\,dz' - \int_0^\pi \left\{G(\mu)-G\left(\frac{z'+\pi}{m}\right)\right\}\frac{\sin z'}{z'+\pi}\,dz'$$
$$+\int_0^\pi\left\{G(\mu)-G\left(\frac{z'+2\pi}{m}\right)\right\}\frac{\sin z'}{z'+2\pi}\,dz' - \ldots\ldots;$$

and on the right-hand side, each integral is positive and less than the preceding one, the signs being alternate; and such a series cannot vanish.

The number ξ depends on m, and on the function $G(z)$; it may happen that as m is increased indefinitely, ξ diminishes indefinitely in such a way that $m\xi$ has a finite limit a. Whether this happens or not, we see from (3), that $\int_0^\mu G(z)\frac{\sin mz}{z}dz$ does not numerically exceed $\pi G(\mu)$, and μ may be so chosen that this is less than the arbitrarily chosen positive number $\frac{1}{2}\epsilon$. Let a fixed value of μ be taken so that this condition is satisfied.

Since $\int_\mu^{\xi_1}\frac{\sin mz}{\sin z}dz$, $\int_{\xi_1}^{\frac{\pi}{2}}\frac{\sin mz}{\sin z}dz$ both converge to zero, as m is indefinitely increased, even though ξ_1 in general depends on m, since, by (2), each integral is numerically less than $\frac{4}{m}\operatorname{cosec}\mu$, we now see that m_1 can be so fixed that

$$\int_0^{\frac{\pi}{2}}F(z)\frac{\sin mz}{\sin z}dz-\frac{\pi}{2}F(+0)$$

is numerically $<\epsilon$, if $m \geqq m_1$; and therefore the expression converges to zero, as m is indefinitely increased. Thus the theorem has been established upon which the proof of Dirichlet's theorem depends, in the generalized form for the case of a function with limited total fluctuation.

UNIFORM CONVERGENCE OF FOURIER'S SERIES.

451. It is known that the limiting sum of a series of continuous functions of a variable is non-uniformly convergent in the neighbourhood of a point of discontinuity of the sum-function, but that the sum-function is not necessarily uniformly convergent in an interval in which it is continuous. In the case of Fourier's series it can be shewn that, if the function $f(x)$ be of limited total fluctuation in its domain $(-\pi, \pi)$, then the series converges uniformly in its whole domain, provided $f(x)$ is continuous in its domain, and provided also $f(-\pi+0)=f(\pi-0)$, so that the function obtained by extending $f(x)$ beyond the range $(-\pi, \pi)$, in accordance with the periodic law $f(x)=f(x+2\pi)$, is also continuous at the points $-\pi, +\pi$. It can be shewn more generally that, provided $f(x)$ is of limited total fluctuation in $(-\pi, \pi)$, and is continuous in an interval (a, b) contained in $(-\pi, \pi)$, so that (a, b) contains no discontinuity of the function in its interior or at its ends, then the series converges uniformly in (a, b). If $f(-\pi+0)$, $f(\pi-0)$ are unequal, the points $-\pi$, π must be included among the points of discontinuity, and therefore cannot be end-points of (a, b). It has been shewn in § 450, that

$$\left|\int_0^{\frac{\pi}{2}} F(z)\frac{\sin mz}{\sin z}\,dz - \frac{\pi}{2}F(+0)\right| < \pi G(\mu) + \frac{4}{m\sin\mu}[F(\mu+0) - F(+0)]$$
$$+ \frac{4}{m\sin\xi_1}\left[F\left(\frac{\pi}{2}-0\right) - F(+0)\right]$$
$$< \pi G(\mu) + \frac{4}{m\sin\mu}\left\{[F(\mu+0) - F(+0)] + F\left(\frac{\pi}{2}-0\right) - F(+0)\right\}.$$

Using this inequality, and the corresponding one for the function $F(-z)$, we have, at any point of (a, b),

$$S_{2m+1} - f(x) < \left(\mu\,\mathrm{cosec}\,\mu + \frac{4}{m\pi}\,\mathrm{cosec}\,\mu\right)[\,|f(x+2\mu) - f(x)|$$
$$+ |f(x-2\mu) - f(x)|\,]$$
$$+ \frac{4}{m\pi}\,\mathrm{cosec}\,\mu\,[\,|f(x+\pi) - f(x)| + |f(x-\pi) - f(x)|\,]$$
$$< \mu\,\mathrm{cosec}\,\mu\,[\,|f(x+2\mu) - f(x)| + |f(x-2\mu) - f(x)|\,] + \frac{A}{m}\,\mathrm{cosec}\,\mu,$$

where A is some fixed number, independent of m, and depending on the upper limit of $|f(x)|$ in the whole interval $(-\pi, \pi)$.

Since $f(x)$ is continuous in (a, b), and is, by the theorem of § 175, therefore also uniformly continuous, a value μ_1 of μ can be chosen such that, for every value of x in (a, b), $|f(x+2\mu) - f(x)|$, $|f(x-2\mu) - f(x)|$ are less than an arbitrarily prescribed positive number ϵ, provided $\mu \leqq \mu_1$. Also a value μ_2 of μ may be so chosen that $\epsilon\mu_2\,\mathrm{cosec}\,\mu_2 < \frac{1}{2}\eta$, where η is an arbitrarily assigned positive number. Take for μ the lesser of the values μ_1, μ_2; then we see that

$$|S_{2m+1} - f(x)| < \eta + \frac{A}{m}\,\mathrm{cosec}\,\mu,$$

for every value of x in (a, b). It follows that, since η and m are independent of x, for all values of n greater than some fixed value n_1, $|S_{2m+1} - f(x)| < \delta$, for the whole interval (a, b), where δ is an arbitrarily chosen number, and n_1 depends only on δ. It has now been shewn that S_{2m+1} converges uniformly to $f(x)$ in the interval (a, b). The function $F(z)$ has been assumed to be monotone; for the general case we have only to consider the difference of two such functions.

The following theorem has now been established:—

In the case of a function $f(x)$ with limited total fluctuation in $(-\pi, \pi)$, the corresponding Fourier's series converges to the value $f(x)$ uniformly in any interval which contains, neither in its interior nor at an end-point, any point of discontinuity of the function.

It must be remembered that the point π is to be reckoned as a point of discontinuity, unless the conditions $f(\pi-0) = f(\pi) = f(-\pi+0)$ are satisfied. A similar remark applies to the point $-\pi$.

THE LIMITING VALUES OF THE COEFFICIENTS IN FOURIER'S SERIES.

452. If the function $f(x)$ be limited, and satisfy Dirichlet's conditions, or in the more general case in which $f(x)$ is a function with limited total fluctuation in the interval $(-\pi, \pi)$, an estimate may be found for the upper limit of the general coefficient in the Fourier's series, as n increases indefinitely.

Let $f(x) = f_1(x) - f_2(x)$, where $f_1(x)$, $f_2(x)$ are monotone functions; then

$$\int_{-\pi}^{\pi} f_1(x)\cos nx\,dx = f_1(-\pi+0)\int_{-\pi}^{\alpha}\cos nx\,dx + f_1(\pi-0)\int_{\alpha}^{\pi}\cos nx\,dx$$
$$= \frac{1}{n}\{f_1(-\pi+0)\sin n\alpha - f_1(\pi-0)\sin n\alpha\};$$

therefore $$\left|\int_{-\pi}^{\pi} f_1(x)\cos nx\,dx\right| \leqq \left|\frac{1}{n}\{f_1(-\pi+0) - f_1(\pi-0)\}\right|$$

where α is some number between $-\pi$ and π. Since a similar result holds for $\int_{-\pi}^{\pi} f_2(x)\cos nx\,dx$, we see that the coefficient $\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,dx$ is numerically less than $\frac{A}{n}$, where A is some fixed number. In a similar manner it can be shewn that the coefficient of $\sin nx$ is numerically less than $\frac{B}{n}$, where B is some fixed number.

Since a series, of which $\frac{1}{n}$ is the general term, is divergent, it is seen that the convergence of Fourier's series is in general not absolute, but depends upon the variation of sign in the terms.

Next, let us suppose that near the point $x = \beta$, where $-\pi \leqq \beta \leqq \pi$, the function $f(x)$ has indefinitely great values, and is such that, near the point β, $f(x) = \frac{\phi(x)}{(x-\beta)^\nu}$, where $\nu < 1$, and $\phi(x)$ is limited, and has only a finite number of oscillations, in the neighbourhood of the point β. In this case the integral $\int_{-\pi}^{\pi} f(x)\cos nx\,dx$ is an improper integral which is represented by the sum of

$$\int_{-\pi}^{\beta-\epsilon}\frac{\phi(x)}{(x-\beta)^\nu}\cos nx\,dx,\quad \int_{\beta+\epsilon}^{\pi}\frac{\phi(x)}{(x-\beta)^\nu}\cos nx\,dx,$$

where ϵ has the limit zero. We may consider the portion

$$\int_{\beta+\epsilon}^{\alpha}\frac{\phi(x)}{(x-\beta)^\nu}\cos nx\,dx$$

of one of these integrals, where α is so chosen that $\phi(x)$ is monotone in the interval $(\beta+\epsilon, \alpha)$. This integral is equivalent to

$$\phi(\alpha')\int_{\beta+\epsilon}^{\alpha}\frac{\cos nx}{(x-\beta)^\nu}dx, \text{ or to } \phi(\alpha')\int_{\epsilon}^{\alpha-\beta}\frac{\cos n(y+\beta)}{y^\nu}dy,$$

where α' lies between $\beta+\epsilon$ and α, and $x=y+\beta$. Let $ny=z$, then

$$\int_{\epsilon}^{\alpha-\beta}\frac{\cos n(y+\beta)}{y^\nu}dy = n^{\nu-1}\int_{\epsilon n}^{n(\alpha-\beta)}\left\{\frac{\cos z}{z^\nu}\cos n\beta - \frac{\sin z}{z^\nu}\sin n\beta\right\}dz.$$

Now the two integrals

$$\int_{\epsilon n}^{n(\alpha-\beta)}\frac{\cos z}{z^\nu}dz, \quad \int_{\epsilon n}^{n(\alpha-\beta)}\frac{\sin z}{z^\nu}dz$$

are in absolute value less than fixed numbers, whatever values ϵ, n may have; for the two integrals

$$\int_0^\infty \frac{\cos z}{z^\nu}dz, \quad \int_0^\infty \frac{\sin z}{z^\nu}dz$$

converge to the values

$$\frac{\pi}{2\cos\frac{1}{2}\nu\pi\,.\,\Gamma(\nu)}, \quad \frac{\pi}{2\sin\frac{1}{2}\nu\pi\,.\,\Gamma(\nu)} \text{ respectively.}$$

It may be proved in a similar manner that the integral

$$\int_{-\pi}^{\beta-\epsilon}\frac{\phi(x)}{(x-\beta)^\nu}\cos nx\,dx$$

contains a portion of the same character as that already obtained. It thus appears* that the coefficient of $\cos nx$ in the Fourier's series is less in absolute value than $A/n^{1-\nu}$, where A is some fixed number; and a corresponding result holds as regards the coefficient of $\sin nx$.

453. The following more general theorem will now be established†:—

If $f(x)$ be a function which is integrable in an interval (a, b), the integral being either a Riemann integral, or an absolutely convergent improper integral, in accordance with the definition of de la Vallée-Poussin, or of Harnack, then $\int_a^b f(x)\sin nx\,dx$, and $\int_a^b f(x)\cos nx\,dx$ both converge to the limit zero, as the positive number n (either integral or not) is indefinitely increased.

All those points of discontinuity of $f(x)$ at which the saltus is $\geqq$ a fixed positive number k may be enclosed in the interiors of the intervals of a finite set $\{\delta\}$, of which the sum is arbitrarily small. Let (α, β) denote one of the intervals of the finite set which is complementary to the set $\{\delta\}$. Since $f(x)$

* See Heine's *Kugelfunctionen*, vol. I, p. 62.

† This theorem is a generalization of a theorem given by Stäckel, nearly equivalent in form to that case of the above which arises when $f(x)$ is continuous in the interval. See *Leipzig. Ber.*, vol. LIII, 1901, also *Nouvelles Annales*, ser. 4, vol. II, 1902.

is absolutely integrable in (a, b), the set $\{\delta\}$ may be so chosen that the sum of the integrals of $|f(x)|$ taken through the set $\{\delta\}$ may be less than an arbitrarily chosen number ζ. It follows that $\int f(x)\sin nx\,dx$ taken through the set $\{\delta\}$ is numerically less than ζ, and therefore $\int_a^b f(x)\sin nx\,dx$ differs, for each value of n, from $\Sigma\int_\alpha^\beta f(x)\sin nx\,dx$, by less than ζ; the summation referring to all the intervals such as (α, β). If ϵ be an arbitrarily chosen positive number, the interval (α, β) can, in accordance with the theorem of § 185, be divided into a number r, of equal parts, such that the fluctuation of $f(x)$ is, in each part, less than $k + \epsilon$.

We have now

$$\int_\alpha^\beta f(x)\sin nx\,dx = \sum_{s=0}^{s=r-1} f\left(\alpha + \frac{s\,\overline{\beta-\alpha}}{r}\right)\int_{\alpha+\frac{s}{r}(\beta-\alpha)}^{\alpha+\frac{s+1}{r}(\beta-\alpha)} \sin nx\,dx$$

$$+\sum_{s=0}^{s=r-1}\int_{\alpha+\frac{s}{r}(\beta-\alpha)}^{\alpha+\frac{s+1}{r}(\beta-\alpha)}\left[f(x) - f\left(\alpha + \frac{s\,\overline{\beta-\alpha}}{r}\right)\right]\sin nx\,dx,$$

where $$\left|f(x) - f\left(\alpha + \frac{s\,\overline{\beta-\alpha}}{r}\right)\right| < k + \epsilon, \text{ for each sub-interval.}$$

From the above identity, we now deduce that

$$\left|\int_\alpha^\beta f(x)\sin nx\,dx\right| < \frac{2r}{n}U + (k+\epsilon)(\beta - \alpha),$$

where U may be taken to denote the upper limit of $|f(x)|$ in all the intervals such as (α, β).

By addition of this result as applied to all the intervals (α, β), we have

$$\left|\Sigma\int_\alpha^\beta f(x)\sin nx\,dx\right| < \frac{2t}{n}U + (k+\epsilon)(b-a)$$

where t denotes the sum of the values of r which correspond to the different intervals.

It now follows that

$$\left|\int_a^b f(x)\sin nx\,dx\right| < \zeta + \frac{2t}{n}U + (k+\epsilon)(b-a);$$

and therefore $$\overline{\lim_{n=\infty}}\left|\int_a^b f(x)\sin nx\,dx\right| \leqq \zeta + (k+\epsilon)(b-a).$$

Since ζ, ϵ, k are all arbitrarily small, it has therefore been proved that

$$\lim_{n=\infty}\int_a^b f(x)\sin nx\,dx = 0.$$

The proof is precisely similar, in case $\cos nx$ be substituted for $\sin nx$.

454. The above theorem is a particular case of the following theorem due to Lebesgue*:—

If $f(x)$ be any function, either limited or unlimited which has a Lebesgue integral in the interval (a, b), then $\int_a^b f(x)\sin nx\,dx$, and $\int_a^b f(x)\cos nx\,dx$, converge to the limit zero, as the positive number n (not necessarily integral) is indefinitely increased. The theorem also holds if the interval (a, b) be replaced by any measurable set of points.

First, let $f(x)$ be a limited function. The interval (a, b) can be divided into a finite number of measurable sets, in each of which the fluctuation of $f(x)$ is less than an arbitrarily chosen positive number ϵ. Let $f_1(x)$ be a function which is constant in each of the sets, and is equal to one of the values of $f(x)$ in that set. We have, since $|f_1(x)-f(x)| < \epsilon$,

$$\left|\int_a^b f(x)\sin nx\,dx - \int_a^b f_1(x)\sin nx\,dx\right| < \epsilon(b-a).$$

Let $e_1, e_2, \ldots e_p$ denote the sets of points in (a, b), for which $f_1(x)$ has the values $c_1, c_2, \ldots c_p$; then

$$\int_a^b f_1(x)\sin nx\,dx = \sum_{q=1}^{q=p} c_q \int_{e_q} \sin nx\,dx.$$

It has been shewn in § 83, that the set e_q is contained in a finite, or infinite, set of intervals of measure $m(e_q)+\eta$, where η is an arbitrarily chosen positive number; it follows that $\int \sin nx\,dx$ taken through e_q, differs from the value of $\int \sin nx\,dx$ taken through the set of intervals which includes e_q, by less than η. Moreover $\int \sin nx\,dx$ taken through this finite, or infinite, set of intervals, converges to zero, as n is indefinitely increased. For, if the intervals be arranged in descending order of magnitude, the integral is numerically less than $\zeta + 2r/n$, where the integer r is such that the sum of all intervals after the first r is less than the arbitrarily chosen positive number ζ; therefore, for a sufficiently large value of n, the integral is numerically $< 2\zeta$, which is arbitrarily small. It follows that

$$\overline{\lim_{n=\infty}} \left|\int_a^b f(x)\sin nx\,dx\right| \leqq \epsilon(b-a) + \eta \sum_1^p c_q;$$

keeping ϵ, and therefore p, fixed, η may be diminished indefinitely, and therefore

$$\overline{\lim_{n=\infty}} \left|\int_a^b f(x)\sin nx\,dx\right| \leqq \epsilon(b-a).$$

* *Annales sc. de l'école normale, supérieure*, ser. 3, vol. xx, 1903. The theorem is stated by Lebesgue for the case $a=0$, $b=2\pi$, of the Fourier's coefficients corresponding to $f(x)$.

Since ϵ is arbitrarily small, it follows that $\int_a^b f(x) \sin nx\, dx$ converges to zero.

Precisely the same argument would apply if the interval (a, b) were replaced by any measurable set of points.

Next, let the function $f(x)$ be unlimited; then a set E exists, containing all those points of (a, b) for which

$$|f(x)| > N;$$

and the number N can be chosen so great that

$$\int_E |f(x)|\, dx < \epsilon,$$

where ϵ is arbitrarily small. If E_1 be the set of points complementary to E, we have

$$\int_a^b f(x) \sin nx dx = \int_E f(x) \sin nx dx + \int_{E_1} f(x) \sin nx dx;$$

and it has been proved above that

$$\int_{E_1} f(x) \sin nx dx$$

converges to zero, as n is indefinitely increased. It follows that

$$\overline{\lim_{n=\infty}} \left| \int_a^b f(x) \sin nx dx \right| < \epsilon,$$

and, since ϵ is arbitrarily small, that

$$\lim_{n=\infty} \int_a^b f(x) \sin nx dx = 0.$$

The substitution of $\cos nx$ for $\sin nx$ makes no essential difference in the proof. As before, the substitution of any measurable set of points for the interval (a, b) makes no essential difference in the proof.

It will be observed that the theorem does not apply to the case in which $f(x)$ has only a non-absolutely convergent improper integral in (a, b).

If we let $a = -\pi$, $b = +\pi$, we obtain the following theorem:—

The coefficients of $\cos nx$, $\sin nx$ in any Fourier's series whatever converge to zero, when n is indefinitely increased.

This is a property of the coefficients which is entirely independent of the convergence or non-convergence of the Fourier's series.

It does not necessarily hold for the generalized Fourier's series, as defined in § 439.

455. The following method for estimating an upper limit to the value of the integral

$$\int_\alpha^\beta \phi(z)\frac{\sin mz}{\sin z}\,dz,$$

where
$$0<\alpha<\beta\leqq\frac{\pi}{2},$$

is due to Schwarz*, and furnishes an alternative proof that the integral converges to zero, when m is indefinitely increased. It is applicable to the case in which $\phi(z)$ is limited, and has a Riemann integral in the interval (α,β).

Let the interval (α,β) be divided into the parts

$$\left(\alpha,\frac{p\pi}{m}\right),\ \left(\frac{p\pi}{m},\frac{\overline{p+1}\,\pi}{m}\right)\dots\left(\frac{q\pi}{m},\beta\right),$$

where p, q are positive integers so chosen that

$$\frac{(p-1)\,\pi}{m}<\alpha\leqq\frac{p\pi}{m},\quad\frac{q\pi}{m}<\beta\leqq\frac{\overline{q+1}\,\pi}{m}.$$

In the first interval let

$$\phi(z)=\phi\left(\frac{p\pi}{m}\right)+\psi_1(z);$$

in the second interval let

$$\phi(z)=\phi\left(\frac{p\pi}{m}\right)+\psi_2(z);$$

in the third interval let

$$\phi(z)=\phi\left(\frac{\overline{p+2}\,\pi}{m}\right)+\psi_3(z);$$

in the fourth interval let

$$\phi(z)=\phi\left(\frac{\overline{p+2}\,\pi}{m}\right)+\psi_4(z),$$

and so on. The given integral I may be written in the form

$$I=\phi\left(\frac{p\pi}{m}\right)\left[\int_\alpha^{\frac{p\pi}{m}}\frac{\sin mz}{\sin z}\,dz+\int_{\frac{p\pi}{m}}^{\frac{\overline{p+1}\pi}{m}}\frac{\sin mz}{\sin z}\,dz\right]$$

$$+\phi\left(\frac{\overline{p+2}\,\pi}{m}\right)\left[\int_{\frac{\overline{p+1}\pi}{m}}^{\frac{\overline{p+2}\pi}{m}}\frac{\sin mz}{\sin z}\,dz+\int_{\frac{\overline{p+2}\pi}{m}}^{\frac{\overline{p+3}\pi}{m}}\frac{\sin mz}{\sin z}\,dz\right]$$

$$+\dots\dots\dots$$

$$+\int_\alpha^\beta\psi(z)\frac{\sin mz}{\sin z}\,dz.$$

* See Sachs' "Versuch zur Geschichte der Darstellung willkürlicher Functionen." *Schlömilch's Zeitschrift*, supplementary vol. xxv, 1880.

If we use the notation of § 445, this may be written in the form

$$I = \phi\left(\frac{p\pi}{m}\right)\left\{\int_{\alpha}^{\frac{p\pi}{m}} \frac{\sin mz}{\sin z}\,dz + (-1)^p \rho_p\right\}$$
$$+ \phi\left(\frac{\overline{p+2}\pi}{m}\right)(-1)^{p+1}(\rho_{p+1} - \rho_{p+2}) + \dots$$
$$+ \int_{\alpha}^{\beta} \psi(z)\frac{\sin mz}{\sin z}\,dz,$$

or

$$I = \phi\left(\frac{p\pi}{m}\right)\left\{\int_{\alpha}^{\frac{p\pi}{m}} \frac{\sin mz}{\sin z}\,dz - \int_{\frac{\overline{p-1}\pi}{m}}^{\frac{p\pi}{m}} \frac{\sin mz}{\sin \alpha}\,dz\right\}$$
$$+ (-1)^{p-1}\left[\phi\left(\frac{p\pi}{m}\right)\left(\frac{2}{m\sin\alpha} - \rho_p\right)\right.$$
$$\left. + \phi\left(\frac{\overline{p+2}\pi}{m}\right)(\rho_{p+1} - \rho_{p+2}) + \dots\right] + \int_{\alpha}^{\beta}\psi(z)\frac{\sin mz}{\sin z}\,dz.$$

Now

$$\int_{\alpha}^{\frac{p\pi}{m}} \frac{\sin mz}{\sin z}\,dz - \int_{\frac{\overline{p-1}\pi}{m}}^{\frac{p\pi}{m}} \frac{\sin mz}{\sin \alpha}\,dz$$

is numerically less than

$$\frac{1}{\sin\alpha}\int_{\frac{\overline{p-1}\pi}{m}}^{\alpha} \sin mz\,dz,$$

or than $\frac{2}{m}$ cosec α, which is less than $\frac{\pi}{m\alpha}$. Also $\rho_p < \frac{2}{m\sin\alpha} < \frac{\pi}{m\alpha}$, and $\rho_r > \rho_{r+1}$; hence we have

$$\left| I - \int_{\alpha}^{\beta}\psi(z)\frac{\sin mz}{\sin z}\,dz\right| < \frac{\pi c}{m\alpha} + c\left(\frac{\pi}{m\alpha} - \rho_p + \rho_{p+1} - \rho_{p+2} + \dots\right) < \frac{2\pi c}{m\alpha},$$

where c is the greatest of the numbers $\left|\phi\left(\frac{p\pi}{m}\right)\right|$, $\left|\phi\left(\frac{\overline{p+2}\pi}{m}\right)\right|$....

Again, $$|\psi_r(z)\sin mz| \leqq \sigma_r(z),$$

where $\sigma_r(z)$ denotes the fluctuation of $\phi(z)$ in the rth interval; hence

$$\left|\int_{\alpha}^{\beta}\psi(z)\frac{\sin mz}{\sin z}\,dz\right| \leqq \int_{\alpha}^{\beta}\frac{\sigma(z)}{\sin z}\,dz \leqq \frac{\pi}{2}\int_{\alpha}^{\beta}\frac{\sigma(z)}{z}\,dz,$$

where $\sigma(z)$, in any interval

$$\left(\frac{r\pi}{m}, \frac{\overline{r+1}\pi}{m}\right),$$

is equal to the fluctuation of $\phi(z)$ in that interval.

We have now obtained Schwarz's theorem, that *the value of the integral*

$$\int_\alpha^\beta \phi(z)\frac{\sin mz}{\sin z}\,dz, \textit{ for } 0<\alpha<\beta\leqq \tfrac{1}{2}\pi,$$

is numerically less than

$$\frac{2c\pi}{m\alpha}+\frac{\pi}{2}\int_\alpha^\beta \frac{\sigma(z)}{z};$$

c denoting the upper limit of $|\phi(z)|$ *in the interval* (α, β), *and* $\sigma(z)$ *the fluctuation of* $\phi(z)$ *in the interval*

$$\left(\frac{r\pi}{m}, \frac{\overline{r+1}\,\pi}{m}\right),$$

and having this constant value throughout that interval.

The function $\phi(z)$ being finite and integrable, in accordance with Riemann's definition, we have

$$\int_\alpha^\beta \frac{\sigma(z)}{z}\,dz<\frac{1}{\alpha}\int_\alpha^\beta \sigma(z)\,dz<\frac{1}{\alpha}\,\Sigma\sigma_r(z)\,\Delta_r,$$

where Δ_r is the length of the rth interval. The sum $\Sigma\sigma_r(z)\,\Delta_r$ becomes, in accordance with Riemann's condition of integrability, arbitrarily small, by making m large enough; and thus we see that

$$\int_\alpha^\beta \phi(z)\frac{\sin mz}{\sin z}\,dz$$

has the limit zero, when m is indefinitely increased.

SUFFICIENT CONDITIONS OF CONVERGENCE OF FOURIER'S SERIES AT A POINT.

456. The function $f(x)$ will be throughout assumed to be integrable in accordance with Lebesgue's definition, through every finite interval, so that the corresponding Fourier's series exists, whether it be convergent or not. It has been shewn in § 444, that, at a point x, the sum of the first $2n+1$ terms of the series is expressed by

$$S_{2n+1}=\frac{1}{\pi}\int_0^{\frac{1}{2}\pi}[f(x+2z)+f(x-2z)]\frac{\sin(2n+1)z}{\sin z}\,dz.$$

The function $f(x+2z)+f(x-2z)$ being integrable in the interval $(0, \frac{1}{2}\pi)$, of z, $\frac{1}{\sin z}[f(x+2z)+f(x-2z)]$ is certainly integrable in any interval, contained in $(0, \frac{1}{2}\pi)$, which does not contain the point $z=0$, at which $\sin z$ vanishes. It follows therefore, by applying the general theorem of § 454, that

$$\frac{1}{\pi}\int_\mu^{\frac{1}{2}\pi}[f(x+2z)+f(x-2z)]\frac{\sin(2n+1)z}{\sin z}\,dz$$

converges to zero, as n is indefinitely increased; where μ is any fixed number such that $0 < \mu \leqq \frac{1}{2}\pi$. The investigation of the limiting value of S_{m+1} is therefore reduced to that of

$$\frac{1}{\pi}\int_0^{\mu} [f(x + 2z) + f(x - 2z)] \frac{\sin(2n+1)z}{\sin z},$$

where μ is arbitrarily small, and subject to the condition $0 < \mu \leqq \frac{1}{2}\pi$. For a given value of x, this integral involves only the functional values of $f(x)$ in the neighbourhood $(x - 2\mu, x + 2\mu)$ of the point x; and consequently the convergence of the series, at this given point x, depends only on the nature of the function in the arbitrarily small neighbourhood of that point. The following theorem has thus been established:—

If $f(x)$ be any function, limited or not, which has a Lebesgue integral in the interval $(-\pi, \pi)$, the convergence of the corresponding Fourier's series at any particular point depends only on the nature of the function $f(x)$ in the arbitrarily small neighbourhood of that point.

This theorem was first established* by Riemann, for the case of a function integrable in accordance with his definition. The theorem of Dirichlet and its extension, which have been investigated above, contain sufficient conditions of convergence of the series in the whole interval $(-\pi, \pi)$. It appears however from the preceding theorem that the convergence or non-convergence of the series at a particular point depends only on the nature of the function in an arbitrarily small neighbourhood of that particular point, and is independent of the general character of the function in the whole interval $(-\pi, \pi)$, this character being limited only by the condition that $f(x)$ must be integrable through the interval, in accordance with Lebesgue's definition, so that the Fourier's coefficients exist as absolutely convergent integrals. Various conditions will accordingly be found, the fulfilment of any one of which is sufficient to ensure the convergence of the series at a particular point.

457. The theorem that $\int_\alpha^\beta F(z) \frac{\sin mz}{\sin z} dz$, where $0 < \alpha < \beta \leqq \frac{1}{2}\pi$, converges to the limit zero, when m is indefinitely increased, was deduced in § 446, as a corollary from the theorem that $\int_0^\alpha F(z) \frac{\sin mz}{\sin z} dz$ converges to the limit $\frac{1}{2}\pi F(+0)$, in the case in which $F(z)$ is a function with limited total fluctuation. It has now however been shewn that the convergence of $\int_\alpha^\beta F(z) \frac{\sin mz}{\sin z} dz$ is really independent of the condition that the function should be one of limited total fluctuation. Referring to the investigation in § 451, it is now clear that it is sufficient for the convergence of

$$\int_0^{\frac{1}{2}\pi} F(z) \frac{\sin mz}{\sin z} dz, \text{ to the limit } \tfrac{1}{2}\pi F(+0),$$

* See his memoir "Ueber die Darstellbarkeit," *Math. Werke*, p. 227.

that $F(z)$ should be of limited total fluctuation in the arbitrarily small interval $(0, \mu)$, of z. For the condition of convergence at x_1 is unaffected by alteration of the values of $f(x)$ outside the interval $(x_1 - 2\mu, x_1 + 2\mu)$. We thus obtain the following sufficient condition of convergence of the Fourier's series at a point x:—

I. *That the Fourier's series may converge at a point x, to the value $\frac{1}{2}\{f(x+0)+f(x-0)\}$, it is sufficient that a neighbourhood of x can be determined, so small that the function $f(x)$ is of limited total fluctuation in that neighbourhood.*

As a particular case of this theorem we have the following:—

I a. *That the Fourier's series may converge at a point x, to the value $\frac{1}{2}\{f(x+0)+f(x-0)\}$, it is sufficient that a neighbourhood $(x-\mu, x+\mu)$ can be determined, such that the function $f(x)$ is monotone, both in $(x-\mu, x)$ and in $(x, x+\mu)$.*

Since $\dfrac{1}{\pi}\displaystyle\int_0^{\frac{1}{2}\pi} \frac{\sin(2n+1)z}{\sin z}\,dz$ converges to the value $\frac{1}{2}$, the condition of convergence of the series at a point x, is satisfied if

$$\int_0^{\frac{1}{2}\pi} \{f(x+2z)-f(x+0)\}\frac{\sin mz}{\sin z}\,dz$$

and

$$\int_0^{\frac{1}{2}\pi} \{f(x-2z)-f(x-0)\}\frac{\sin mz}{\sin z}\,dz$$

each converge to the limit zero, as m is indefinitely increased. It is however not necessary that these conditions be satisfied; all that is necessary is that the sum

$$\int_0^{\frac{1}{2}\pi}[f(x+2z)+f(x-2z)-\lim_{t=0}\{f(x+t)-f(x-t)\}]\frac{\sin mz}{\sin z}\,dz$$

should converge to zero. This remark enables us to obtain conditions of convergence of greater generality than those which would be obtained by assuming that each of the two separate parts of this integral converges to zero.

At a point x, of continuity of $f(x)$, the necessary condition is that

$$\int_0^{\frac{1}{2}\pi} \{f(x+2z)+f(x-2z)-2f(x)\}\frac{\sin mz}{\sin z}\,dz$$

should converge to the limit zero.

Writing, for convenience

$$2z = t,\quad f(x+t)+f(x-t)=\phi(t),$$

the condition of convergence of the series is that

$$\int_0^{\pi}\frac{\phi(t)-\phi(+0)}{t}\,\frac{t}{\sin\frac{1}{2}t}\sin\tfrac{1}{2}mt\,dt$$

should converge to zero, as m is indefinitely increased.

Now
$$\frac{\phi(t)-\phi(+0)}{t}\cdot\frac{t}{\sin\frac{1}{2}t}$$
is certainly integrable in any interval, contained in $(0, \pi)$, which does not contain the point $t = 0$. In case
$$\left|\frac{\phi(t)-\phi(+0)}{t}\right|$$
is integrable in the interval $(0, \mu)$ of t, so also is
$$\left|\frac{\phi(t)-\phi(0)}{t}\,\frac{t}{\sin\frac{1}{2}t}\right|;$$
and then the theorem of § 454, suffices to shew that
$$\int_0^\pi \frac{\phi(t)-\phi(+0)}{t}\,\frac{t}{\sin\frac{1}{2}t}\sin\tfrac{1}{2}mt\,dt$$
converges to the limit zero.

We have accordingly obtained the following theorem:—

II. *If* $\left|\frac{\phi(t)-\phi(+0)}{t}\right|$ *be integrable in* $(0, \mu)$, *where* $0 < \mu \leqq \pi$, *and* $\phi(t)$ *denotes* $f(x+t)+f(x-t)$, *then the Fourier's series is convergent at the point* x. *This condition is satisfied when* $f(x+0)$, $f(x-0)$ *are both definite, and* $\left|\frac{f(x+t)-f(x+0)}{t}\right|$, $\left|\frac{f(x-t)-f(x-0)}{t}\right|$ *are both integrable in* $(0, \mu)$; *or else when* $f(x+0)$, $f(x-0)$ *are not necessarily definite, but* $\phi(+0)$ *is so, and* $\left|\frac{\phi(t)-\phi(+0)}{t}\right|$ *is integrable in* $(0, \mu)$.

The series in either case converges to $\frac{1}{2}\phi(+0)$.

II a. *If* x *be a point of continuity of* $f(x)$, *the Fourier's series converges to the value* $f(x)$, *at the point* x, *if* $\left|\frac{f(x+t)+f(x-t)-2f(x)}{t}\right|$ *be integrable in the interval* $(0, \mu)$; *and in particular if*
$$\left|\frac{f(x+t)-f(x)}{t}\right|,\quad \left|\frac{f(x-t)-f(x)}{t}\right|$$
are both integrable in $(0, \mu)$.

The condition that $\left|\frac{\phi(t)-\phi(+0)}{t}\right|$ should be integrable in $(0, \mu)$ is satisfied if $\overline{\lim_{t=0}}\,\frac{\phi(t)-\phi(+0)}{t}$, $\underline{\lim_{t=0}}\,\frac{\phi(t)-\phi(+0)}{t}$ are both finite. In case x be a point of continuity of the function, this condition is satisfied if the four derivatives $D^+f(x)$, $D_+f(x)$, $D^-f(x)$, $D_-f(x)$, at the point x, are all finite, and in particular if $f(x)$ have a finite differential coefficient. We thus obtain the following theorems:—

III. *The Fourier's series converges at a point x, if*

$$\overline{\lim_{t=0}}\frac{\phi(t)-\phi(+0)}{t}, \quad \underline{\lim_{t=0}}\frac{\phi(t)-\phi(+0)}{t}$$

be both finite, $\phi(t)$ denoting $f(x+t)+f(x-t)$. In particular if x be a point at which $f(x)$ has an ordinary discontinuity, and

$$\lim_{t=+0}\frac{f(x+t)-f(x+0)}{t}, \quad \lim_{t=+0}\frac{f(x-t)-f(x-0)}{-t}$$

be both either definite, or indefinite between finite limits of indeterminacy, then the series is convergent at the point x.

III a. *If $f(x)$ be continuous at x, the Fourier's series converges, at the point x, to the value $f(x)$, if the four derivatives of $f(x)$ at the point be finite, and in particular if $f(x)$ have a finite differential coefficient at the point.*

It is however not necessary for the integrability of $\left|\frac{\phi(t)-\phi(+0)}{t}\right|$ in the neighbourhood of $t=0$, that it should be limited in that neighbourhood. It is sufficient (see § 281, Ex. 1) that

$$\left|\frac{\phi(t)-\phi(+0)}{t}\right| \leqq \frac{A}{t^a}, \text{ or that } |\phi(t)-\phi(+0)| \leqq At^{1-a},$$

where $1-a$ is some positive number, and A is some fixed positive number, for all values of t which are not greater than some fixed value μ. We thus obtain the following sufficient conditions of convergence of the series:—

IV. *The Fourier's series converges, at a point x, if, for all values of t not greater than some fixed positive number μ, the condition*

$$|\phi(t)-\phi(+0)| \leqq At^k,$$

be satisfied, A and k denoting fixed positive numbers.

IV a. *The Fourier's series converges to $f(x)$, at a point x of continuity of the function, if $|f(x+t)-f(x)| \leqq At^k$, where A and k are fixed positive numbers, for all values of t numerically less than some fixed positive number μ.*

At a point of ordinary discontinuity, it is sufficient that both

$$|f(x+t)-f(x+0)| \text{ and } |f(x-t)-f(x-0)|$$

should satisfy this condition.

A more general sufficient condition of integrability of $\left|\frac{\phi(t)-\phi(+0)}{t}\right|$ in the neighbourhood of $t=0$, is that, in a sufficiently small interval $(0, \mu)$, of t,

$$|\phi(t)-\phi(+0)| \leqq \frac{A}{\log\frac{1}{t}\log\log\frac{1}{t}\cdots\cdots\left[\log\log\cdots\frac{1}{t}\right]^{1+a}}$$

where A and a are positive numbers (see § 299). We therefore obtain the following condition of convergence of the series:—

V. *The Fourier's series converges, at a point x, to the value*

$$\tfrac{1}{2}\lim_{t=0}\{f(x+t)+f(x-t)\},$$

if, for all positive values of t not exceeding some fixed value μ, the condition

$$|\phi(t)-\phi(+0)| \leqq \frac{A}{\log\frac{1}{t}\log\log\frac{1}{t}\cdots\cdots\left[\log\log\ldots\frac{1}{t}\right]^{1+a}}$$

be satisfied, where A and a are fixed positive numbers. In particular it is sufficient that both

$$|f(x+t)-f(x+0)|,\quad |f(x-t)-f(x-0)|$$

should satisfy this condition.

CONDITIONS OF UNIFORM CONVERGENCE OF FOURIER'S SERIES.

458. Conditions will now be investigated that the Fourier's series, corresponding to a summable function $f(x)$, either limited in $(-\pi, \pi)$, or unlimited but integrable, may be uniformly convergent in a given interval contained in $(-\pi, \pi)$. With a view to this investigation, the following general theorem* will be established:—

The function $f(x)$ being summable and integrable in $(-\pi, \pi)$, each of the four integrals $\int_a^\beta f(x\pm 2z)\chi(z)\begin{smallmatrix}\sin\\ \cos\end{smallmatrix} mz\,dz$, taken through any interval (a, β), such that $0 \leqq a < \beta \leqq \frac{1}{2}\pi$, converges to the limit zero, as the positive number m is indefinitely increased, uniformly for all values of x contained in the interval $(-\pi, \pi)$; the function $\chi(z)$ being any function with limited total fluctuation in (a, β). The function $f(x)$ is assumed to be such that

$$f(x\pm 2\pi)=f(x),\ \textit{for}\ -\pi<x<\pi.$$

More generally, $\sin mz$ or $\cos mz$ may be replaced by $\phi(mz)$, where $\phi(z)$ is any limited summable function, of which the integral, taken through any finite interval whatever, is less, in absolute magnitude, than some fixed positive number, independent of the particular interval. It is unnecessary that m be restricted to be integral.

First, it will be assumed that $f(x)$ is a limited function. It is sufficient to consider the case of the integral $\int_a^\beta f(x+2z)\chi(z)\sin mz\,dz$; the proof in the case of the other three integrals being precisely similar. Also, the substitution of $\phi(mz)$ for $\sin mz$ makes no essential difference in the proof.

* Hobson, "On the uniform convergence of Fourier's series," *Proc. Lond. Math. Soc.*, ser. 2, vol. v.

Let U and L denote the upper, and the lower, limit of $f(x+2z)$, for all values of x in the interval $(-\pi, \pi)$, and for all values of z in the interval (α, β). Let the interval (L, U) be divided into p portions (c_0, c_1), $(c_1, c_2) \dots (c_{q-1}, c_q) \dots (c_{p-1}, c_p)$, where $c_0 = L$, $c_p = U$, and where $c_q - c_{q-1} < \epsilon$, for every value of q. Let the function $f_1(x+2z)$ be defined as follows:—For those values of $x+2z$ for which $c_0 \leqq f(x+2z) < c_1$, let $f_1(x+2z) = c_0$; for those values of $x+2z$ for which $c_1 \leqq f(x+2z) < c_2$, let $f_1(x+2z) = c_1$; and generally, let $f_1(x+2z) = c_{q-1}$, when $c_{q-1} \leqq f(x+2z) < c_q$; when $f(x+2z) = c_p$, let $f_1(x+2z) = c_p$. For any particular value of x, it may, for example, happen that there are no values of z such that $c_0 \leqq f(x+2z) < c_1$; in that case there are no values of $x+2z$, with the given value of x, for which $f_1(x+2z) = c_0$. We have

$$\left| \int_\alpha^\beta f(x+2z)\,\chi(z)\sin mz\,dz - \int_\alpha^\beta f_1(x+2z)\,\chi(z)\sin mz\,dz \right| < \epsilon(\beta-\alpha)\,\bar{\chi},$$

where $\bar{\chi}$ is the upper limit of $|\chi(z)|$ in the interval (α, β); and this holds for all values of m, and of x in $(-\pi, \pi)$.

We have also

$$\int_\alpha^\beta f_1(x+2z)\,\chi(z)\sin mz\,dz = \sum_{q=0}^{q=p} c_q \int_{e_q} \chi(z)\sin mz\,dz,$$

where e_q is that set of points z, at which $c_q \leqq f(x+2z) < c_{q-1}$; this set e_q depending upon the value of x.

In the interval $(-2\pi, 2\pi)$ of the variable x, let E_q denote that set of points at each of which $c_q \leqq f(x) < c_{q+1}$. Let the set E_q be enclosed in a finite, or enumerably infinite, set H_q of non-overlapping intervals, such that $m(H_q) - m(E_q) = \eta$. For any fixed value of x, the set e_q consists of that part of E_q which lies in the interval $(x+2\alpha, x+2\beta)$, contained in $(-2\pi, 2\pi)$. Let that part of the set of intervals E_q which lies in $(x+2\alpha, x+2\beta)$ be denoted by F_q; then it can be shewn that $m(F_q) - m(e_q) \leqq \eta$. For, if possible, let $m(F_q) - m(e_q) = \eta + \gamma$, where γ is a positive number. Let the set e_q be enclosed in the interiors of non-overlapping intervals of a set L_q, all in the interval $(x+2\alpha, x+2\beta)$, such that $m(L_q) < m(e_q) + \gamma$; and let $\bar{H}_q$ denote that set of intervals which consists of L_q together with that part of H_q which is not in $(x+2\alpha, x+2\beta)$. We have then

$$\begin{aligned} m(\bar{H}_q) &= m(H_q) + m(L_q) - m(F_q) \\ &< m(H_q) - \eta \\ &< m(E_q). \end{aligned}$$

As E_q cannot be enclosed in a set of intervals $\bar{H}_q$, of measure less than $m(E_q)$, it is impossible that the positive number γ can exist; and therefore

$$m(F_q) - m(e_q) \leqq \eta.$$

It is to be observed that the number η is independent of the parameter x, in the integral.

We have now

$$\left|\int_{e_q} \chi(z)\sin mz\,dz - \int_{F_q} \chi(z)\sin mz\,dz\right| < \bar{\chi}\,\eta.$$

Let the intervals of the set H_q, in descending order of length, be denoted by $\gamma_1, \gamma_2, \gamma_3 \ldots\ldots$ In case the point $x+2\alpha$, or $x+2\beta$, is interior to an interval of H_q, we divide that interval into two parts, and assign separate indices to those parts. We may choose r so that

$$m(H_q) - (\gamma_1 + \gamma_2 + \ldots + \gamma_r) < \eta.$$

Of the intervals $\gamma_1, \gamma_2, \ldots \gamma_r \ldots\ldots$ let those which fall in $(x+2\alpha, x+2\beta)$ be $\gamma_{s_1}, \gamma_{s_2}, \gamma_{s_3}, \ldots\ldots$, where $s_1 < s_2 < s_3 \ldots$; and let s_t be the greatest of these indices which does not exceed r. We have then

$$\gamma_{s_t+1} + \gamma_{s_t+1} + \ldots < \eta, \quad \text{and} \quad m(F_q) - (\gamma_{s_1} + \gamma_{s_2} + \ldots + \gamma_{s_t}) < \eta;$$

or denoting by D_q the finite set of intervals $\gamma_{s_1}, \gamma_{s_2}, \ldots \gamma_{s_t}$, we have

$$m(F_q) - m(D_q) < \eta.$$

The number t, of the intervals in the set D_q, cannot exceed the number r, which is independent of the value of x.

We now have

$$\left|\int_{F_q} \chi(z)\sin mz\,dz - \int_{D_q} \chi(z)\sin mz\,dz\right| < \eta\bar{\chi}.$$

The function $\chi(z)$ having limited total fluctuation in the interval (α, β), it may be expressed as the difference $\chi_1(z) - \chi_2(z)$, of two functions $\chi_1(z)$, $\chi_2(z)$, each of which is monotone in (α, β).

The integral $\int_\mu^\lambda \chi(z)\sin mz\,dz$ may, by means of the second mean value theorem, be expressed as

$$\chi_1(\mu)\int_\mu^\xi \sin mz\,dz + \chi_1(\lambda)\int_\xi^\lambda \sin mz\,dz - \chi_2(\mu)\int_\mu^{\xi'} \sin mz\,dz - \chi_2(\lambda)\int_{\xi'}^\lambda \sin mz\,dz,$$

where ξ, ξ' are two numbers in the interval (μ, λ). We thus see that

$$\left|\int_\mu^\lambda \chi(z)\sin mz\,dz\right| < \frac{4}{m}(\bar{\chi}_1 + \bar{\chi}_2),$$

where $\bar{\chi}_1$, $\bar{\chi}_2$ are the upper limits of $|\chi_1(z)|$ and of $|\chi_2(z)|$ in (α, β); the interval (μ, λ) being supposed to be contained in (α, β). We have now

$$\left|\int_{D_q} \chi(z)\sin mz\,dz\right| < \frac{4}{m}(\bar{\chi}_1 + \bar{\chi}_2)\,t < \frac{4}{m}(\bar{\chi}_1 + \bar{\chi}_2)\,r,$$

since t cannot exceed r.

By combining the inequalities which have been obtained, we find that

$$\left|\int_\alpha^\beta f(x+2z)\,\chi(z)\sin mz\,dz\right| < \epsilon(\beta-\alpha)\,\bar{\chi} + \left(2\bar{\chi}\eta + \frac{4r}{m}\bar{\chi}'\right)\sum_{q=0}^{q=p} c_q,$$

where $\bar{\chi}'$ denotes $\bar{\chi}_1 + \bar{\chi}_2$; and this holds for every value of x in $(-\pi, \pi)$. Let a positive number ζ be now arbitrarily chosen; we can then choose ϵ so that $\epsilon(\beta-\alpha)\,\bar{\chi} < \frac{1}{3}\zeta$. Having fixed ϵ accordingly, and consequently also the numbers $c_0, c_1, \ldots c_q$ being capable of being fixed, we next choose η so that $2\bar{\chi}\eta\sum_{q=0}^{q=p} c_q < \frac{1}{3}\zeta$; the number r is then fixed. We can now choose a value m_1 of m, such that $\frac{4r}{m}\bar{\chi}'\sum_{q=0}^{q=p} c_q < \frac{1}{3}\zeta$, for $m \geqq m_1$.

It has now been shewn that, having given a positive number ζ, arbitrarily small, a number m_1 can be so determined that

$$\left|\int_\alpha^\beta f(x+2z)\,\chi(z)\sin mz\,dz\right| < \zeta,$$

for $m \geqq m_1$, and for all values of x in the interval $(-\pi, \pi)$. It has therefore been shewn that, when $f(x)$ is a limited summable function,

$$\int_\alpha^\beta f(x+2z)\,\chi(z)\sin mz\,dz$$

converges to the limit zero, as m is indefinitely increased, uniformly for all values of x in the interval $(-\pi, \pi)$; and consequently also for all values of x in any interval (a, b) contained in $(-\pi, \pi)$.

Next, let $f(x)$ be no longer limited, but still integrable in accordance with Lebesgue's definition.

If ζ be an arbitrarily fixed positive number, a positive number N can be so determined that

$$\int |f(x)|\,dx < \tfrac{1}{2}\zeta/\bar{\chi},$$

the integral being taken over that set of points K_N in the interval $(-2\pi, 2\pi)$, for each of which $|f(x)| > N$. If k_N be that part of K_N which lies in the interval $(x+2\alpha, x+2\beta)$, for any fixed value of x belonging to the interval $(-\pi, \pi)$, we have, *a fortiori*

$$\int_{k_N} |f(x)|\,dx < \tfrac{1}{2}\zeta/\bar{\chi}.$$

Let the function $f_2(x+2z)$ be defined by the specifications

$$f_2(x+2z) = f(x+2z), \quad \text{if} \quad |f(x+2z)| \leqq N;$$

and

$$f_2(x+2z) = 0, \quad \text{if} \quad |f(x+2z)| > N.$$

Thus $f_2(x+2z)$ vanishes at all the points of the set K_N; and it is a limited summable function.

We have now

$$\int_\alpha^\beta f(x+2z)\,\chi(z)\sin mz\,dz$$
$$=\int_{kN} f(x+2z)\,\chi(z)\sin mz\,dz+\int_\alpha^\beta f_2(x+2z)\,\chi(z)\sin mz\,dz.$$

By the first part of the theorem, we see that a value m_1 of m can be determined so that

$$\left|\int_\alpha^\beta f_2(x+2z)\,\chi(z)\sin mz\,dz\right|<\tfrac{1}{2}\zeta,$$

for $m \geqq m_1$, and for all values of x in $(-\pi, \pi)$.

Also $$\left|\int_{kN} f(x+2z)\,\chi(z)\sin mz\,dz\right|<\tfrac{1}{2}\zeta;$$

hence we have shewn that

$$\left|\int_\alpha^\beta f(x+2z)\,\chi(z)\sin mz\,dz\right|<\zeta,$$

provided $m \geqq m_1$, for all values of x in $(-\pi, \pi)$. The theorem has therefore been completely established.

459. Some particular cases of the general theorem established in § 458 will now be considered.

(1) Let $\alpha=0$, $\beta=\frac{1}{2}\pi$, and let x have the single value 0; also let $\chi(z)=1$, and $m=2n$. Then, changing z into $\frac{1}{2}x$, we see that

$$\int_0^\pi f(x)\sin nx\,dx,\quad \int_0^\pi f(x)\cos nx\,dx$$

both converge to zero, as n is indefinitely increased. Taking the integral which involves $f(x-2z)$, we see that

$$\int_{-\pi}^0 f(x)\sin nx\,dx,\quad \int_{-\pi}^0 f(x)\cos nx\,dx$$

also converge to zero. By addition, we obtain the theorem already established in § 454, that the Fourier's coefficients

$$\int_{-\pi}^\pi f(x)\,{}^{\sin}_{\cos}\,nx\,dx$$

converge to zero, as n is indefinitely increased.

(2) Let $0<\alpha$, $\beta=\frac{1}{2}\pi$, and $\chi(z)=\operatorname{cosec} z$, which is of limited fluctuation in the interval $(\alpha, \frac{1}{2}\pi)$. We see then that, if (a, b) be an interval for x, in which $f(x)$ is limited, and be also such that $f(a-0)$, $f(b+0)$ are finite, then

$$\int_\alpha^{\frac{1}{2}\pi}\left[f(x+2z)+f(x-2z)-f(x+0)-f(x-0)\right]\frac{\sin mz}{\sin z}\,dz$$

converges uniformly to zero in the interval (a, b), of x.

For, by the theorem, the two integrals

$$\int_\alpha^{\frac{1}{2}\pi} f(x+2z)\frac{\sin mz}{\sin z}\,dz, \quad \int_\alpha^{\frac{1}{2}\pi} f(x-2z)\frac{\sin mz}{\sin z}\,dz$$

converge uniformly to zero in (a, b); also $|f(x+0)+f(x-0)|$ is less, for all values of x in (a, b), than some fixed finite number, and

$$\int_\alpha^{\frac{1}{2}\pi} \frac{\sin mz}{\sin z}\,dz$$

converges to zero, as m is indefinitely increased. It thus appears that, in order that a Fourier's series may converge uniformly to zero, in an interval (a, b) contained in $(-\pi, \pi)$, it is sufficient that

$$\int_0^\alpha [f(x+2z)+f(x-2z)-2f(x)]\frac{\sin mz}{\sin z}\,dz$$

should converge to zero, as m is indefinitely increased, uniformly for all values of x in (a, b). We know that it is a necessary condition for such uniform convergence that $f(x)$ should be continuous in (a, b), including the end-points a and b.

In accordance with this result, it depends only upon the nature of the function $f(x)$ in the interval $(a-2\alpha, b+2\alpha)$, where α is arbitrarily small, whether the Fourier's series converge uniformly in (a, b) or not; the nature of $f(x)$ in the remainder of $(-\pi, \pi)$ being irrelevant, subject only to the restriction that $f(x)$ must have a Lebesgue integral in $(-\pi, \pi)$, whether it be limited or not. We have therefore obtained the following theorem:—

If (a, b) be any interval contained in $(-\pi, \pi)$, such that $f(x)$ is continuous in (a, b), including the end-points a and b, then the answer to the question whether the Fourier's series converges uniformly in (a, b), or not, depends only upon the nature of $f(x)$ in an interval (a', b') including (a, b) in its interior, and exceeding it in length by an arbitrarily small amount. The function $f(x)$ may be of any character in the part of $(-\pi, \pi)$ outside (a', b'), so long as it has a Lebesgue integral in $(-\pi, \pi)$.

This theorem contains, for the theory of uniform convergence, the parallel to the theorem of § 456, that convergence or non-convergence of the series, at a particular point x, depends only on the nature of the function in an arbitrarily small neighbourhood of x. The latter theorem is that particular case of the theorem here established, which arises when the interval (a, b) is reduced to a particular point x.

(3) It has been shewn in § 451, that, if $f(x)$ be a function with limited total fluctuation in $(-\pi, \pi)$, the Fourier's series corresponding to $f(x)$ converges uniformly in any interval (a, b) which contains no point of discontinuity of the function, either in its interior or at its ends. By applying the theorem obtained in (2), we now see that the following extension of this result holds:—

The function $f(x)$ being summable and integrable, whether it be limited or not; if (a', b') be any interval, contained in $(-\pi, \pi)$, and such that $f(x)$ is of limited total fluctuation in (a', b'), then the Fourier's series, corresponding to $f(x)$, converges uniformly in any interval (a, b) in the interior of (a', b'), provided the function be continuous in (a, b), including the points a and b.

(4) Let the function $\chi(z)$ be defined by $\chi(0) = 0$, and

$$\chi(z) = \frac{1}{z} - \frac{1}{\sin z}, \text{ for } z > 0;$$

also let $\alpha = 0$, $\beta = \mu < \frac{1}{2}\pi$. We then see that

$$\int_0^{\mu} [f(x+2z) + f(x-2z) - 2f(x)] \left(\frac{1}{z} - \frac{1}{\sin z}\right) \sin mz \, dz$$

converges uniformly to zero, as m is indefinitely increased, in any interval (a, b) in which $f(x)$ is limited.

It thus appears that, if

$$\int_0^{\mu} [f(x+2z) + f(x-2z) - 2f(x)] \frac{\sin mz}{\sin z} dz$$

converges uniformly in (a, b), then so also does

$$\int_0^{\mu} [f(x+2z) + f(x-2z) - 2f(x)] \frac{\sin mz}{z} dz.$$

Therefore, *the condition of uniform convergence of the series in an interval (a, b), in which $f(x)$ is continuous, including the points a and b, is that*

$$\int_0^{\mu} [f(x+2z) + f(x-2z) - 2f(x)] \frac{\sin mz}{z} dz$$

should converge uniformly to zero in the interval, as m is indefinitely increased.

The number μ may here be taken arbitrarily small; in fact, in accordance with the theorem of § 458, if $0 < \mu_1 < \mu$, the integral

$$\int_{\mu_1}^{\mu} [f(x+2z) + f(x-2z) - 2f(x)] \frac{\sin mz}{z} dz$$

converges to zero, as m is indefinitely increased, uniformly in (a, b).

460. We proceed to apply the result in (4), of § 459, to obtain sufficient conditions for the uniform convergence of the series in an interval (a, b).

Denoting $\quad f(x+2z) + f(x-2z) - 2f(x)$, by $F(z)$,

we have, if $\quad 0 < \mu_1 < \mu$,

$$\left| \int_0^{\mu} \frac{F(z)}{z} \sin mz \, dz \right| \leq \left| \int_0^{\mu_1} \frac{F(z)}{z} \sin mz \, dz \right| + \left| \int_{\mu_1}^{\mu} \frac{F(z)}{z} \sin mz \, dz \right|.$$

Let it now be assumed that, for every value of x in (a, b), the integral $\int_0^{\mu_1} \frac{F(z)}{z} dz$ exists as a Lebesgue integral, and that $\int_0^{\mu_1} \left| \frac{F(z)}{z} \right| dz$ converges to the limit zero, as μ_1 is indefinitely diminished, uniformly for all values of x in (a, b).

We have then

$$\left|\int_0^{\mu_1} \frac{F(z)}{z} \sin mz\, dz\right| < \int_0^{\mu_1} \left|\frac{F(z)}{z}\right| dz.$$

The number μ_1 can now be chosen so small that, if ζ be an arbitrarily fixed positive number, the inequality

$$\int_0^{\mu_1} \left|\frac{F(z)}{z}\right| dz < \tfrac{1}{2}\zeta$$

is satisfied for this value of μ_1, and for every value of x in (a, b). The number μ_1 having been so fixed, we can fix a value m_1, of m, such that

$$\left|\int_{\mu_1}^{\mu} \frac{F(z)}{z} \sin mz\, dz\right| < \tfrac{1}{2}\zeta,$$

for $m \geqq m_1$, and for every value of x in (a, b). We have then

$$\left|\int_0^{\mu} \frac{F(z)}{z} \sin mz\, dz\right| < \zeta,$$

for $m \geqq m_1$, and for every value of x in (a, b).

The following theorem has therefore been established:—

It is a sufficient condition for the uniform convergence of the Fourier's series, in an interval (a, b) in which $f(x)$ is continuous, the points a and b included, that

$$\int_0^{\mu_1} \left|\frac{f(x+2z) + f(x-2z) - 2f(x)}{z}\right| dz$$

should exist for all values of x in (a, b), and should converge to zero, as μ_1 is indefinitely diminished, uniformly for all values of x in (a, b). The condition is satisfied if the two integrals

$$\int_0^{\mu_1} \left|\frac{f(x+2z) - f(x)}{z}\right| dz, \quad \int_0^{\mu_1} \left|\frac{f(x-2z) - f(x)}{z}\right| dz$$

both exist, and are uniformly convergent.

In particular, the series is uniformly convergent in any interval (a, b) in which one of the four derivatives, and therefore each of the other three derivatives, is limited, the end-points a and b being included. A special case is that in which $f(x)$ has a limited differential coefficient in (a, b), including a and b.

The condition is also satisfied if

$$|f(x \pm \beta) - f(x)| < C\beta^k,$$

for all values of x in (a, b), and for all positive values of β not greater than some fixed positive number; where C and k are positive numbers independent of x. This is a generalization of the sufficient condition, obtained in § 457, for the convergence of the series at a particular point.

461. Let the integral $\int_0^{\mu} \frac{F(z)}{z} \sin mz dz$ be expressed in the form

$$\left[\int_0^{\frac{2\pi}{m}} + \int_{\frac{2\pi}{m}}^{\frac{3\pi}{m}} + \int_{\frac{3\pi}{m}}^{\frac{4\pi}{m}} + \ldots + \int_{\frac{2r\pi}{m}}^{\mu}\right] \frac{F(z)}{z} \sin mz dz,$$

where r is an integer such that

$$0 \leqq \mu - \frac{2r\pi}{m} < \frac{2\pi}{m}.$$

We assume that (a, b) is contained in an interval (a', b'), in which $f(x)$ is limited; if then we choose μ to be less than the smaller of the numbers

$$\tfrac{1}{2}(a - a'), \quad \tfrac{1}{2}(b' - b),$$

we see that $F(z)$ is limited, for all values of z in $(0, \mu)$, and for all values of x in (a, b). We have now

$$\left|\int_0^{\frac{2\pi}{m}} F(z) \frac{\sin mz}{z} dz\right| < m \int_0^{\frac{2\pi}{m}} |F(z)| dz$$

$$< 2\pi \times \text{upper limit of } |F(z)| \text{ in } \left(0, \frac{2\pi}{m}\right).$$

Since a continuous function is uniformly continuous, the two functions

$$f(x + 2z) - f(x), \quad f(x - 2z) - f(x)$$

converge to zero, as z converges to zero, uniformly for all values of x in (a, b). It follows that

$$\left|\int_0^{\frac{2\pi}{m}} F(z) \frac{\sin mz}{z} dz\right| < \eta_m$$

where η_m converges to zero, as m is indefinitely increased, and is independent of the value of x.

Next, we have

$$\left|\int_{\frac{2r\pi}{m}}^{\mu} F(z) \frac{\sin mz}{z} dz\right| < \frac{4/m}{\mu - 2\pi/m} \times \text{upper limit of } |F(z)| \text{ in } \left(\frac{2r\pi}{m}, \mu\right)$$

$$< \frac{4/m}{\mu - 2\pi/m} \times \text{upper limit of } |F(z)| \text{ in } (0, \mu)$$

$$< \eta_m',$$

where η_m' converges to zero, as m is indefinitely increased, and is independent of the value of x.

The remaining part of the integral may be written in the form

$$\int_0^{\frac{\pi}{m}} \left[\frac{F\left(z+\frac{2\pi}{m}\right)}{z+\frac{2\pi}{m}} - \frac{F\left(z+\frac{3\pi}{m}\right)}{z+\frac{3\pi}{m}} + \frac{F\left(z+\frac{4\pi}{m}\right)}{z+\frac{4\pi}{m}} - \dots + \frac{F\left(z+\frac{\overline{2r-2}\,\pi}{m}\right)}{z+\frac{\overline{2r-2}\,\pi}{m}} - \frac{F\left(z+\frac{\overline{2r-1}\,\pi}{m}\right)}{z+\frac{\overline{2r-1}\,\pi}{m}} \right] \sin mz\,dz,$$

which is less, in absolute value, than

$$\int_0^{\pi} \left| \frac{F\left(\frac{z+2\pi}{m}\right)}{z+2\pi} - \frac{F\left(\frac{z+3\pi}{m}\right)}{z+3\pi} + \dots - \frac{F\left(\frac{z+\overline{2r-1}\,\pi}{m}\right)}{z+\overline{2r-1}\,\pi} \right| dz;$$

and this does not exceed

$$\int_0^{\pi} \sum_{s=1}^{s=r-1} \left| \frac{F\left(\frac{z+2s\pi}{m}\right)}{z+2s\pi} - \frac{F\left(\frac{z+\overline{2s+1}\,\pi}{m}\right)}{z+\overline{2s+1}\,\pi} \right| dz.$$

Now

$$\frac{F\left(\frac{z+2s\pi}{m}\right)}{z+2s\pi} - \frac{F\left(\frac{z+\overline{2s+1}\,\pi}{m}\right)}{z+\overline{2s+1}\,\pi} = \frac{F\left(\frac{z+2s\pi}{m}\right) - F\left(\frac{z+\overline{2s+1}\,\pi}{m}\right)}{z+\overline{2s+1}\,\pi} + \pi \frac{F\left(\frac{z+2s\pi}{m}\right)}{(z+2s\pi)(z+\overline{2s+1}\,\pi)};$$

hence

$$\left| \frac{F\left(\frac{z+2s\pi}{m}\right)}{z+2s\pi} - \frac{F\left(\frac{z+\overline{2s+1}\,\pi}{m}\right)}{z+\overline{2s+1}\,\pi} \right| \leqq \frac{\left| F\left(\frac{z+2s\pi}{m}\right) - F\left(\frac{z+\overline{2s+1}\,\pi}{m}\right) \right|}{(2s+1)\pi} + \frac{1}{2s(2s+1)\pi} \left| F\left(\frac{z+2s\pi}{m}\right) \right|.$$

We now see that the part of the integral to be estimated is, in absolute value, less than

$$\left(\frac{1}{3}+\frac{1}{5}+\dots+\frac{1}{2r-1}\right)\Delta + \frac{1}{2}\,.\,\text{upper limit of } |F(z)| \text{ in } (0, \mu);$$

where Δ is the greatest of the numbers

$$\left| F\left(\frac{z+2s\pi}{m}\right) - F\left(\frac{z+\overline{2s+1}\,\pi}{m}\right) \right|,$$

for $s = 1, 2, 3, \dots r-1$, and for all values of z in $(0, \pi)$.

The upper limit of $|F(z)|$, or $|f(x+2z)+f(x-2z)-2f(x)|$, for the interval $(0, \mu)$, of z, and for all values of x in (a, b), depends upon μ, and is a number $2u(\mu)$, which may be made as small as we please by taking μ sufficiently small.

Also

$$\left(\frac{1}{3}+\frac{1}{5}+\dots+\frac{1}{2r-1}\right)\Delta<\left(1+\frac{1}{2}+\frac{1}{3}+\dots+\frac{1}{2r}\right)\Delta<(C_r+\log 2r)\,\Delta,$$

where C_r is a number between 0 and 1, which converges to Mascheroni's constant. Let D_m denote the greatest value of the difference of the values of $f(x)$ at the ends of an interval of length π/m, contained in the interval $(a-\mu,\ b+\mu)$, for all possible positions of such sub-interval; thus $\Delta \leqq D_m$.

It has now been proved that, for $0<\mu<\frac{1}{2}\pi$,

$$\left|\int_0^\mu [f(x+2z)+f(x-2z)-2f(x)]\frac{\sin mz}{z}\,dz\right| < \eta_m+\eta_m'+u(\mu)$$
$$+D_m\left\{C_r+\log\frac{m}{\pi}+\log\left(\mu-\frac{2\theta\pi}{m}\right)\right\},$$

where θ is less than unity, and such that $\mu-\dfrac{2r\pi}{m}=\theta\,.\,\dfrac{2\pi}{m}$.

We have now to find a sufficient condition that it be possible, with μ fixed, to determine a value $\overline{m}$ of m, corresponding to an arbitrarily prescribed positive number ζ, such that

$$\left|\int_0^\mu F(z)\,\frac{\sin mz}{m}\,dz\right|<\zeta,\ \text{for}\ m\geqq\overline{m},$$

and for all values of x in (a, b).

If μ_1 be a number such that $0<\mu_1<\mu$, we can choose μ_1 so small that

$$u(\mu_1)<\tfrac{1}{8}\zeta.$$

The number μ_1 having been so chosen, we can now choose a value m_1, of m, such that

$$\left|\int_{\mu_1}^\mu F(z)\,\frac{\sin mz}{z}\,dz\right|<\tfrac{1}{8}\zeta,\ \text{for}\ m\geqq m_1,$$

and for all values of x in (a, b); this follows from the uniform convergence of the integral to the limit zero.

We then have

$$\left|\int_0^\mu F(z)\frac{\sin mz}{z}\,dz\right|<\tfrac{1}{8}\zeta+\eta_m+\eta'_m+D_m\left\{C_r+\log\frac{m}{\pi}+\log\left(\mu_1-\frac{2\theta_1\pi}{m}\right)\right\},$$

the numbers η'_m, D_m, θ_1 now having the values of the former η'_m, D_m, θ which correspond to μ_1 instead of μ. We can now choose m_2 so that $\eta_m<\frac{1}{8}\zeta$, for $m\geqq m_2$; also we can choose m_3 so that $\eta'_m<\frac{1}{8}\zeta$, for $m\geqq m_3$. Again, since

D_m converges to zero, as m is indefinitely increased, the function $f(x)$ being assumed to be continuous in the interval $(a-2\mu_1,\ b+2\mu_1)$, we can so determine m_4 that

$$D_m\left\{C_r+\log\left(\mu_1-\frac{2\theta_1\pi}{m}\right)\right\}<\tfrac{1}{8}\zeta,\ \text{for } m\geqq m_4.$$

Let us now assume that it is possible so to choose m_5, that

$$D_m\log\frac{m}{\pi}<\tfrac{1}{8}\zeta,\ \text{for } m\geqq m_5.$$

Taking $\bar{m}$ to be the greatest of the numbers m_1, m_2, m_3, m_4, m_5, we now have

$$\left|\int_0^{\mu}F(z)\frac{\sin mz}{z}\,dz\right|<\zeta,\ \text{for } m\geqq\bar{m}.$$

Therefore, with the assumption made, that $D_m\log\frac{m}{\pi}$ converges to zero, as m is indefinitely increased, it has been shewn that the convergence of the Fourier's series in (a, b) is uniform. The following theorem has now been established:—

If (a', b') be an interval such that, for every pair of points x, $x+\beta$, contained in that interval, $\{f(x\pm\beta)-f(x)\}\log\beta$ converges to zero, as β is indefinitely diminished, uniformly for all values of x, then the Fourier's series converges uniformly in any interval (a, b) contained in the interior of (a', b'). More generally, it is sufficient that $\{f(x+\beta)+f(x-\beta)-2f(x)\}\log\beta$ should satisfy the similar condition.

For the case in which the interval (a, b) is reduced to a single point, the condition of the theorem becomes a sufficient condition of convergence of the Fourier's series at that point. Thus a sufficient condition of convergence of the series at a point is, that a neighbourhood of the point can be determined, such that, if ϵ be any prescribed positive number, another positive number β can be determined, such that

$$|\{f(x\pm\beta)-f(x)\}\log\beta|<\epsilon,$$

for every pair of points x, $x+\beta$ in that neighbourhood. This condition was given by Dini*.

The condition of the theorem is satisfied, in particular, if, for every pair of points x, $x+\beta$, in (a', b'), when β is sufficiently small,

$$|f(x\pm\beta)-f(x)|<C\beta^k,$$

where C, k are positive numbers independent of x. In this form, the condition was given by Lipschitz†, as a sufficient condition of the convergence of the

* *Série di Fourier*, p. 49.

† *Crelle's Journal*, vol. LXIII, p. 308.

Fourier's series at a single point, the condition being applied to a neighbourhood of the point.

The condition of uniform convergence in the interval (a, b), stated in the theorem, is satisfied, if β can be so determined that, for all pairs of points x, $x+\beta$, in the interval (a', b'),

$$|f(x\pm\beta)-f(x)| < \frac{C}{\frac{1}{\beta}\log\frac{1}{\beta}\cdot\log\log\frac{1}{\beta}\cdots\left\{\log\log\cdots\frac{1}{\beta}\right\}^{1+k}}$$

where C, k are positive numbers, independent of x.

FURTHER INVESTIGATIONS OF DIRICHLET'S INTEGRAL.

462. Many investigations relating to Dirichlet's integral have had as their object the determination of sufficient conditions for the convergence of a Fourier's series at a particular point. These investigations have resulted in the discovery of many special conditions sufficient for the convergence of the series, some of them of great generality. The most important of these conditions have been already discussed in the present chapter.

An account will now be given of some investigations of Dirichlet's integral, given* by Kronecker, Hölder, and Brodén†. The mode of procedure of the last writer will be here adopted.

It has been shewn that the question of the convergence of the series at a point depends upon whether $\int_0^\epsilon F(z)\frac{\sin mz}{z}dz\,(0<\epsilon\leqq\frac{1}{2}\pi)$ converges to the limit zero, as m is indefinitely diminished; the function $F(z)$ being such that $F(+0)=0$.

The integral $\int_0^\epsilon F(z)\frac{\sin mz}{z}dz$ may be expressed as the sum

$$\int_0^{\chi(m)} F(z)\frac{\sin mz}{z}dz+\int_{\chi(m)}^{\epsilon} F(z)\frac{\sin mz}{z}dz,$$

where $\chi(m)$ is a function of m, which is positive and $<\epsilon$, for every value of m, and is such that it converges to zero, as m is indefinitely increased; *i.e.*

$$\lim_{m=\infty}\chi(m)=0.$$

Let us consider first the integral

$$I_\chi=\int_0^{\chi(m)} F(z)\frac{\sin mz}{z}dz.$$

* *Berliner Sitzungsber.* 1885, "Ueber das Dirichlet'sche Integral," by Kronecker; and in the same volume, "Ueber eine neue hinreichende Bedingung..." by Hölder.

† *Math. Annalen*, vol. LII.

Since $\left|\frac{\sin mz}{z}\right| < m$, we have $|I_\chi| < m\chi(m)U_\chi$, where U_χ is the upper limit of $|F(z)|$ in the interval $(0, \chi(m))$; and this upper limit U_χ converges to the limit zero, as m is indefinitely increased, because $F(+0) = 0$.

The condition that $\lim\limits_{m=\infty} |I_\chi| = 0$ is that $\chi(m)$ must be such that

$$\lim_{m=\infty} m\chi(m)U_\chi = 0;$$

and this condition will be satisfied if $\chi(m)$ be such that $m\chi(m)$ remains less than some fixed number, however great m may be.

The condition may also be satisfied if $m\chi(m)$ becomes indefinitely great, as m does so, since U_χ may be such that $m\chi(m)U_\chi$ has the limit zero.

In particular, if $\chi(m) = \frac{1}{m}$, we see that

$$\int_0^{\frac{1}{m}} F(z)\frac{\sin mz}{z}\,dz$$

has the limit zero; and therefore

$$\left|\lim_{m=\infty}\int_0^{\chi(m)} F(z)\frac{\sin mz}{z}\,dz\right| = \left|\lim_{m=\infty}\int_{\frac{1}{m}}^{\chi(m)} F(z)\frac{\sin mz}{z}\,dz\right| < U_\chi\,|\log m\chi(m)|.$$

We have now seen that

$$\lim_{m=\infty}\int_0^{\chi(m)} F(z)\frac{\sin mz}{z}\,dz$$

vanishes, provided $\chi(m)$ be such that $\lim\limits_{m=\infty}\chi(m) = 0$, and also such that one at least of the numbers $m\chi(m).U_\chi$, $\log\{m\chi(m)\}\,U_\chi$ has the limit zero; where U_χ is the upper limit of $|F(z)|$ in the interval $(0, \chi(m))$.

If we take $\chi(m) = \frac{\pi\xi}{m}$, where ξ is a positive constant, and change z into $\frac{\pi x}{m}$, then write $\frac{\pi}{m} = \sigma$, we obtain Kronecker's* theorem

$$\lim_{\sigma=0}\int_0^{\xi} F(\sigma x)\frac{\sin \pi x}{x}\,dx = 0, \text{ when } F(+0) = 0.$$

If we choose $\chi(m)$ so that one of the above conditions is satisfied, we have

$$\lim_{m=\infty}\int_0^{\epsilon} F(z)\frac{\sin mz}{z}\,dz = \lim_{m=\infty}\int_{\chi(m)}^{\epsilon} F(z)\frac{\sin mz}{z}\,dz.$$

* *Loc. cit.*, p. 642.

The integral on the right-hand side can be divided into three parts, by dividing the interval $(\chi(m), \epsilon)$ into the sum of the three intervals

$$(\chi(m),\ 2p\pi/m),\quad (2p\pi/m,\ 2q\pi/m),\quad (2q\pi/m,\ \epsilon),$$

where p, q are positive integers, such that

$$0 < \frac{2p\pi}{m} - \chi(m) \leqq \frac{2\pi}{m}, \text{ and } 0 < \epsilon - \frac{2q\pi}{m} \leqq \frac{2\pi}{m}.$$

Of the three integrals, the third clearly has the limit zero; and the first is, in absolute value, less than $F_1 \log \frac{2p\pi}{m\chi(m)}$, where F_1 is the upper limit of $|F(z)|$ in the interval $(0, 2p\pi/m)$. Hence the absolute value of the first integral is less than $F_1 \log\left(1 + \frac{2\pi}{m\chi(m)}\right)$, of which the limit is zero; unless $m\chi(m)$ has the limit zero, in which case the absolute value of the integral is less than

$$F_1 m \left\{\frac{2p\pi}{m} - \chi(m)\right\},$$

or than $2\pi F_1$; and the limit is therefore in this case also zero.

We have now left the integral

$$\int_{\frac{2p\pi}{m}}^{\frac{2q\pi}{m}} F(z) \frac{\sin mz}{z}\, dz.$$

If we divide the interval into portions of length $\frac{2\pi}{m}$, and then change the variable in each portion, we can reduce this integral to the form

$$\int_0^\pi \sin z \sum_{\iota=p}^{\iota=q-1} \left[\frac{F\left(\frac{z+2\iota\pi}{m}\right)}{z+2\iota\pi} - \frac{F\left(\frac{z+\overline{2\iota+1}\pi}{m}\right)}{z+\overline{2\iota+1}\pi}\right] dz,$$

which is less than

$$\int_0^\pi \left| \sum_{\iota=p}^{\iota=q-1} \left[\frac{F\left(\frac{z+2\iota\pi}{m}\right)}{z+2\iota\pi} - \frac{F\left(\frac{z+\overline{2\iota+1}\pi}{m}\right)}{z+\overline{2\iota+1}\pi}\right]\right| dz.$$

It follows that, for a given ϵ, the limit, when m is indefinitely increased, of

$$\int_0^\epsilon F(z) \frac{\sin mz}{z}\, dz$$

is zero, provided it is possible to choose $\chi(m)$, consistently with the conditions which have been laid down, so that, if δ be an arbitrarily small positive number, a value m_1 of m can be found, such that

$$\int_0^\pi \left| \sum_{\iota=p}^{\iota=q-1} \left[\frac{F\left(\frac{z+2\iota\pi}{m}\right)}{z+2\iota\pi} - \frac{F\left(\frac{z+\overline{2\iota+1}\pi}{m}\right)}{z+\overline{2\iota+1}\pi}\right]\right| dz < \delta,$$

for all values of m, such that $m \geqq m_1$. It can easily be verified that the limit of this integral is independent of the value of ϵ, and also of the particular choice of the function $\chi(m)$.

If we take $m\chi(m) = 2n - 1$, where n is a fixed positive integer, and write $z\pi$ for z, and σ for π/m, the sufficient condition takes the form

$$\lim_{\sigma=0} \int_0^1 \left| \sum_{h=2n}^{h=2\left[\frac{\epsilon}{2\sigma}\right]} (-1)^h \frac{F(\sigma z + \sigma h)}{z+h} \right| dz = 0,$$

where $\left[\frac{\epsilon}{2\sigma}\right]$ denotes the integral part of $\frac{\epsilon}{2\sigma}$; this is equivalent to a condition obtained by Kronecker*.

463. The sufficient condition which has been obtained above, that the limit of $\int_0^\epsilon F(z) \frac{\sin mz}{z} dz$ may be zero, is of a very general character, and includes, as special cases, various sufficient conditions which have been obtained by special methods.

The condition will be satisfied if, corresponding to a number δ as small as we please, it is possible to choose $\chi(m)$ so that a value m_1 of m can be found, such that, for $m > m_1$, all those values of z between 0 and π, for which

$$\sum_{\iota=p}^{\iota=q-1} \left\{ \frac{F\left(\frac{z+2\iota\pi}{m}\right)}{z+2\iota\pi} - \frac{F\left(\frac{z+\overline{2\iota+1}\pi}{m}\right)}{z+\overline{2\iota+1}\pi} \right\}$$

numerically exceeds δ, form a set of points of zero content, provided also the absolute value of this sum, for every value of z, has a finite upper limit. For in this case the integral is less than $\pi\delta + \int \Sigma dz$, where the integral is taken over sub-intervals which include all the points at which $|\Sigma| > \delta$, and this is less than

$$\pi\delta + \Sigma_1 \int dz,$$

where Σ_1 is the upper limit of $|\Sigma|$. Since $\int dz$ is, by hypothesis, arbitrarily small, $\pi\delta + \Sigma_1 \int dz$ is arbitrarily small, and thus the limit of the integral is zero.

An important deduction from the general theorem is that

$$\lim_{m=\infty} \int_0^\epsilon F(z) \frac{\sin mz}{z} dz = 0,$$

* *Loc. cit.*, p. 651.

provided the function $F(z)$ is such that

$$\int_0^\epsilon \left|\frac{F(z)}{z}\right| dz$$

has a definite finite value. This theorem has already been established in § 457.

For, we may write the sum

$$\sum_{\iota=p}^{\iota=q-1}\left\{\frac{F\left(\frac{z+2\iota\pi}{m}\right)}{z+2\iota\pi}-\frac{F\left(\frac{z+\overline{2\iota+1}\pi}{m}\right)}{z+\overline{2\iota+1}\pi}\right\}$$

in the form

$$\sum_{\iota=p}^{\iota=q-1}\frac{1}{m}\frac{F\left(\frac{z+2\iota\pi}{m}\right)}{\frac{z+2\iota\pi}{m}}-\sum_{\iota=p}^{\iota=q-1}\frac{1}{m}\frac{F\left(\frac{z+\overline{2\iota+1}\pi}{m}\right)}{\frac{z+\overline{2\iota+1}\pi}{m}};$$

now if $|F(z)/z|$ is integrable in the interval $(0, \epsilon)$, each of the sums in this last expression converges uniformly, as m is increased indefinitely, to the value

$$\int_0^\epsilon \frac{F(z)}{z} dz,$$

and hence the limit converges uniformly, for every value of z in $(0, \pi)$, to zero.

Since $F(z)$ is an integrable function, we know that

$$\int_{\epsilon_1}^{\epsilon_2} \frac{F(z)}{z} dz,$$

where $\epsilon_2 > \epsilon_1 > 0$, has a definite finite value, hence it is immaterial what positive value less than $\frac{1}{2}\pi$, ϵ may have in

$$\int_0^\epsilon \frac{F(z)}{z} dz.$$

The theorem last proved may be used to shew that the two conditions

$$\lim_{m=\infty}\int_0^\epsilon F(z)\frac{\sin mz}{z} dz = 0,$$

$$\lim_{m=\infty}\int_0^\epsilon F(z)\frac{\sin mz}{\sin z} dz = 0$$

are equivalent to one another.

We have

$$\int_0^\epsilon F(z)\frac{\sin mz}{\sin z} dz = \int_0^\epsilon F(z)\frac{\sin mz}{z} dz - \int_0^\epsilon F(z)\, z\, .\, \phi(z)\frac{\sin mz}{z} dz,$$

where $\phi(z)$ denotes $\frac{1}{z}\left(1-\frac{z}{\sin z}\right)$; which for a sufficiently small value of $|z|$ is expansible in a convergent series of ascending powers of z. Since

$$\lim_{z=0} zF(z)\,\phi(z)=0,$$

and $F(z)\,\phi(z)$ is an integrable function, we see, by the last theorem, that

$$\lim_{m=\infty}\int_0^\epsilon F(z)\,.\,z\,\phi(z)\frac{\sin mz}{z}\,dz=0,$$

therefore

$$\lim_{m=\infty}\int_0^\epsilon F(z)\frac{\sin mz}{z}\,dz=\lim_{m=\infty}\int_0^\epsilon F(z)\frac{\sin mz}{\sin z}\,dz.$$

It has thus been shewn that any condition which is sufficient to ensure that

$$\int_0^\epsilon F(z)\frac{\sin mz}{z}\,dz$$

has the limit zero is also sufficient to ensure that

$$\int_0^\epsilon F(z)\frac{\sin mz}{\sin z}\,dz$$

has the limit zero. This has already been established in § 459 (4), in a less restricted class of cases.

464. If ϵ and $\chi(m)$ are such that

$$\left|\sum_{h=2p}^{h=2q-1}(-1)^h F\left(\frac{z+h\pi}{m}\right)\right|$$

is less than a fixed positive number N, for every value of m, it can be shewn that

$$\lim_{m=\infty}\int_0^\epsilon F(z)\frac{\sin mz}{z}\,dz=0.$$

For, by a known arithmetical theorem due to Abel, we have

$$\left|\sum_{\iota=p}^{\iota=q-1}\left\{\frac{F\left(\frac{z+2\iota\pi}{m}\right)}{z+2\iota\pi}-\frac{F\left(\frac{z+\overline{2\iota+1}\pi}{m}\right)}{z+\overline{2\iota+1}\pi}\right\}\right|=\left|\sum_{h=2p}^{h=2q-1}(-1)^h\frac{F\left(\frac{z+h\pi}{m}\right)}{z+h\pi}\right|<\frac{N}{z+2p\pi};$$

where $2p$ is determined by the condition $0<\frac{2p\pi}{m}-\chi(m)\leqq\frac{2\pi}{m}$.

Let $m\chi(m)$ have the fixed positive integral value $2n-1$; then $p=n$, and thus the absolute value of the above sum is less than $\frac{N}{z+2n\pi}$, which may be made arbitrarily small, for $0<z<\epsilon$, by choosing n large enough; and thus the theorem is established.

If the interval $(0,\epsilon)$ can be chosen, such that, in this interval, $F(z)$ is monotone, the condition is satisfied. Suppose $F(z)$ does not diminish as z increases from 0 to ϵ, then all the terms in the sum

$$\sum_{\iota=p}^{\iota=q-1}\left\{F\left(\frac{z+2\iota\pi}{m}\right)-F\left(\frac{z+\overline{2\iota+1}\pi}{m}\right)\right\}$$

are negative or zero, and the numerical value of the sum is less than $|F(\epsilon)|$; and thus the condition is satisfied. This is the case which was considered by Dirichlet, and has been otherwise investigated in § 446. Again

$$\left|F\left(\frac{z+2\iota\pi}{m}\right)-F\left(\frac{z+\overline{2\iota+1}\pi}{m}\right)\right|$$

is, for all the values of z, not greater than the fluctuation of the function $F(z)$ in the interval

$$\left(\frac{2\iota\pi}{m},\ \frac{\overline{2\iota+2}\pi}{m}\right),$$

hence the condition for the vanishing of the limit is satisfied if the total fluctuation of $F(z)$ in the interval $(0, \epsilon)$ is less than a fixed finite number. This is the case which was considered by Jordan, and was given in § 446. The special conditions obtained in § 457 and § 461, due to Dini and Lipschitz, may also be deduced from the general theorem of § 462.

THE NON-CONVERGENCE OF FOURIER'S SERIES.

465. Various sufficient conditions for the convergence of the Fourier's series, corresponding to a given function $f(x)$, have now been investigated. The continuity of $f(x)$ at a particular point x is neither necessary nor sufficient to ensure that the Fourier's series, corresponding to $f(x)$, converges at the point x; it being assumed that the function $f(x)$ is such that the Fourier's series exists. Du Bois Reymond* gave the first example of a Fourier's series, corresponding to a continuous function, which fails to converge at points of a certain everywhere-dense set.

It is not definitely known whether a Fourier's series, corresponding to a continuous function, can be such that the series fails to converge at every point of an interval. It has however been proved† by Fatou, that, in case the coefficients of the series be such that $\lim_{n=\infty} na_n = 0$, $\lim_{n=\infty} nb_n = 0$, the series is convergent at a set of points of which the measure is equal to that of the whole interval $(-\pi, \pi)$.

An example, due to Schwarz‡, will be here given, of a function which is everywhere continuous, but for which the Fourier's series fails to converge at

* *Abhandlungen der bayerischen Akademie*, vol. XII, Abthg. 2.

† *Acta Math.* vol. XXX, p. 379.

‡ See the history of the theory of Fourier's series, by Sachs, *Schlömilch's Zeitschr.* Supplement, vol. XXV.

a certain point. It will here be shewn* that the series is, at that point, in reality, oscillatory. It will then be shewn that the function may be employed to construct another continuous function, for which the Fourier's series fails to converge at each point of an everywhere-dense set.

Let the product $1.3.5 \ldots (2\lambda+1)$ be denoted by $[2\lambda+1]$, and let the function $\phi(z)$ be defined for the interval $(0, \alpha)$, where $0<\alpha \leqq \frac{1}{2}\pi$, in the following manner:—In the interval $(\pi/[\lambda-1], \pi/[\lambda])$, let $\phi(z)=c_\lambda \sin[\lambda]z$, where c_λ is a constant, depending upon the value of λ; let λ have all values $\lambda_1, \lambda_1+1, \lambda_1+2, \ldots$, where λ_1 is a fixed integer, and we may suppose α so chosen that $\alpha=\pi/[\lambda_1-1]$; also let $\phi(0)=0$. If the sequence $c_{\lambda_1}, c_{\lambda_1+1}, c_{\lambda_1+2}, \ldots$ be so chosen that it converges to the limit zero, the function $\phi(z)$ is continuous at the point $z=0$, but it has an indefinitely great number of oscillations in an arbitrarily small neighbourhood of that point. If the constants c_λ satisfy the further condition, that $c_\lambda \log(2\lambda+1)$ becomes indefinitely great, as λ is indefinitely increased, it will be shewn that the integral

$$\int_0^\alpha \phi(z)\frac{\sin(2n+1)z}{z}\,dz$$

will increase indefinitely, as n has successively the values of integers in a certain sequence. Thus the Fourier's series, corresponding to the continuous function defined by $f(x)=0$, for $-\pi \leqq x \leqq 0$, and $f(x)=\phi(\frac{1}{2}x)$, for $0\leqq x\leqq 2\alpha$, and $f(x)=0$, for $2\alpha \leqq x \leqq \pi$, does not converge at the point $x=0$.

Let $2n+1=1.3.5\ldots(2\mu+1)=[\mu]$; then

$$\int_0^\alpha \phi(z)\frac{\sin[\mu]z}{z}\,dz$$

may be written in the form

$$c_\mu\int_{\pi/[\mu]}^{\pi/[\mu-1]}\frac{\sin^2[\mu]z}{z}\,dz+\sum_{r=\lambda_1}^{\mu-1} c_r\int_{\pi/[r]}^{\pi/[r-1]}\frac{\sin[r]z\sin[\mu]z}{z}\,dz$$
$$+\sum_{r=\mu+1}^{\infty} c_r\int_{\pi/[r]}^{\pi/[r-1]}\frac{\sin[r]z\sin[\mu]z}{z}\,dz.$$

The first integral may be written in the form

$$\tfrac{1}{2}c_\mu\int_{\pi/[\mu]}^{\pi/[\mu-1]}\frac{1-\cos 2[\mu]z}{z}\,dz,$$

which is equivalent to

$$\tfrac{1}{2}c_\mu\log(2\mu+1)-\tfrac{1}{2}c_\mu\frac{[\mu]}{\pi}\int_{\pi/[\mu]}^{\beta}\cos 2[\mu]z\,dz,$$

where β is some number between $\pi/[\mu]$ and $\pi/[\mu-1]$.

* See Hobson, "The failure of convergence of Fourier's series," *Proc. Lond. Math. Soc.*, ser. 2, vol. III.

Now let $c_\mu \log(2\mu+1)$ increase indefinitely with μ. This is consistent with c_μ having the limit zero; for we have only to take

$$c_\mu = \{\log(2\mu+1)\}^{-s},$$

where s is some fixed positive number, less than unity.

Since $$c_\mu \frac{[\mu]}{\pi} \int_{\pi/[\mu]}^{\beta} \cos 2[\mu] z \, dz$$

is numerically not greater than c_μ/π, we see that, with the supposition made as to c_μ, the expression

$$c_\mu \int_{\pi/[\mu]}^{\pi/[\mu-1]} \frac{\sin^2[\mu] z}{z} dz$$

becomes indefinitely great, as μ is increased indefinitely.

To evaluate $$\sum_{r=\lambda_1}^{\mu-1} c_r \int_{\pi/[r]}^{\pi/[r-1]} \frac{\sin[r] z \sin[\mu] z}{z} dz,$$

we see, by writing $\sin[r]z \,.\, \sin[\mu]z$ as half the difference of two cosines, and applying the second mean value theorem to each integral, that the absolute value of the expression is less than

$$\sum_{r=\lambda_1}^{\mu-1} c_r \frac{[r]}{\pi} \left\{\frac{1}{[\mu]-[r]} + \frac{1}{[\mu]+[r]}\right\},$$

or than $$\sum_{r=\lambda_1}^{\mu-1} \frac{c_r}{\pi} \frac{[r]}{[\mu-1]} \left\{\frac{1}{2\mu+1-[r]/[\mu-1]} + \frac{1}{2\mu+1+[r]/[\mu-1]}\right\},$$

which is less than

$$\frac{c_{\lambda_1}}{\pi} \Sigma \frac{[r]}{[\mu-1]} \cdot \frac{1}{\mu};$$

and this is less than

$$\frac{c_{\lambda_1}}{\pi\mu} \left\{1 + \frac{1}{2\mu-1} + \frac{1}{(2\mu-1)(2\mu-3)} + \dots\right\}.$$

Therefore the absolute value of the integral is less than $2c_\lambda/\pi\mu$; and this becomes indefinitely small, as μ is indefinitely increased; and therefore the limiting value of the expression is zero.

Lastly, we have to consider the expression

$$\sum_{r=\mu+1}^{\infty} c_r \int_{\pi/[r]}^{\pi/[r-1]} \frac{\sin[r] z \sin[\mu] z}{z} dz.$$

Since $$\left|\frac{\sin[\mu] z}{z}\right| < [\mu], \text{ and } |\sin[r] z| \leq 1,$$

the absolute value of the expression is less than $\pi c_{\mu+1}$; and this has the limit zero, when μ is indefinitely increased.

It has now been shewn that

$$\int_0^a \frac{\phi(z) \sin[\mu] z}{z} dz,$$

increases indefinitely with μ, where $[\mu] = 1.3.5 \ldots (2\mu+1)$, provided c_λ has the value $\{\log(2\lambda+1)\}^{-s}$, where $0 < s < 1$.

466. We proceed to consider the case in which $2n+1 = (2p+1)[\mu-1]$, where p is an integer which varies with μ in such a manner that it always lies between 0 and μ.

In this case, as before, we divide the integral

$$\int_0^a \phi(z)\frac{\sin(2n+1)z}{z}dz$$

into three parts

$$c_\mu \int_{\pi/[\mu]}^{\pi/[\mu-1]} \sin[\mu]z \frac{\sin(2p+1)[\mu-1]z}{z}dz$$
$$+\sum_{r=\lambda_1}^{\mu-1} c_r \int_{\pi/[r]}^{\pi/[r-1]} \frac{\sin[r]z\sin(2p+1)[\mu-1]z}{z}dz$$
$$+\sum_{r=\mu+1}^{\infty} c_r \int_{\pi/[r]}^{\pi/[r-1]} \frac{\sin[r]z\sin(2p+1)[\mu-1]z}{z}dz.$$

The first part is equal to

$$\frac{c_\mu[\mu]}{2\pi}\int_{\pi/[\mu]}^{\beta} [\cos\{[\mu-1](2\mu-2p)z\} - \cos\{[\mu-1](2\mu+2p+2)z\}]dz,$$

where β is a number between $\pi/[\mu]$ and $\pi/[\mu-1]$; and this expression is less, in absolute value, than

$$\frac{c_\mu[\mu]}{\pi}\left\{\frac{1}{[\mu-1](2\mu-2p)} + \frac{1}{[\mu-1](2\mu+2p+2)}\right\},$$

or than

$$\frac{c_\mu}{\pi}\left\{\frac{1+1/2\mu}{1-p/\mu} + \frac{1+1/2\mu}{1+1/\mu+p/\mu}\right\}.$$

If, now, p increases with μ in such a manner that p/μ is always less than some fixed number which is less than unity, then this expression diminishes indefinitely, as μ is indefinitely increased. It would also be sufficient that

$$p/\mu = 1 - \kappa\{\log(2\mu+1)\}^{-s'},$$

where $s' < s$, and $c_\mu = \{\log(2\mu+1)\}^{-s}$; the positive number κ being fixed.

The second part of the above integral is less, in absolute value, than

$$\sum_{r=\lambda_1}^{\mu-1}\frac{c_r[r]}{\pi}\left\{\frac{1}{(2p+1)[\mu-1]-[r]} + \frac{1}{(2p+1)[\mu-1]+[r]}\right\},$$

or than

$$\frac{c_{\lambda_1}}{\pi}\Sigma\frac{[r]}{[\mu-1]}\left\{\frac{1}{2p+1-[r]/[\mu-1]} + \frac{1}{2p+1+[r]/[\mu-1]}\right\},$$

and this is less than

$$\frac{c_{\lambda_1}}{p\pi}\left\{1 + \frac{1}{2\mu-1} + \frac{1}{(2\mu-1)(2\mu-3)} + \ldots\right\},$$

or than $2c_{\lambda_1}/p\pi$. Therefore the expression diminishes indefinitely, as p is indefinitely increased.

That the third part of the above integral has the limit zero is seen from the fact that its absolute value is less than $c_{\mu+1}(2p+1)[\mu-1]\pi/[\mu]$, or than $\pi c_{\mu+1}(2p+1)/(2\mu+1)$.

It has now been proved that

$$\int_0^a \phi(z)\frac{\sin(2n+1)z}{z}dz$$

has the limit zero, if $2n+1$ increases indefinitely through a sequence of the form

$$[\mu_1-1](2p_1+1),\quad [\mu_2-1](2p_2+1),\quad [\mu_3-1](2p_3+1),\ldots$$

where $\mu_1, \mu_2, \mu_3, \ldots$ is an increasing sequence of integers, and $p_1, p_2, p_3, \ldots$ are such that $p/\mu \leqq 1-\kappa\{\log(2\mu+1)\}^{-s'}$, where $s'<s$.

It has now been shewn that the limit of the sum of the Fourier's series oscillates; the limit being infinite, or zero, according as one or other of two particular sequences of values of n is chosen.

467. In order to construct a continuous function which is such that the corresponding Fourier's series fails to converge at all the points of an everywhere-dense set, we take the following definition of $f(x)$:—

If $-\pi \leqq x \leqq \xi$, where ξ is a fixed point in the interval $(-\pi, \pi)$, let $f(x)=0$; if $0 \leqq x-\xi \leqq 2a$, let $f(x)=\phi\left(\frac{x-\xi}{2}\right)$, where $\phi(z)$ is the function that has been already discussed. In case $\xi+2a<\pi$, we take $f(x)=0$, for $\xi+2a<x\leqq\pi$.

The limit of

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x')\frac{\sin\frac{1}{2}(2n+1)(x'-x)}{\sin\frac{1}{2}(x'-x)}dx',$$

at the point ξ, depends upon that of

$$\frac{1}{\pi}\int_0^a \phi(z)\frac{\sin(2n+1)z}{z}dz.$$

It has been shewn that this limit is zero, or is indefinitely great, according as $2n+1$ increases indefinitely through one or other of two sequences. It follows that the Fourier's series for $f(x)$ does not converge at the point ξ.

Let the function $f(x)$ be now denoted by $\psi(x, \xi)$; and let $\xi_1, \xi_2, \ldots \xi_r, \ldots$ be an enumerable set of values of ξ, everywhere-dense in the interval $(-\pi, \pi)$. Let us consider the function

$$F(x)=c_1\psi(x, \xi_1)+c_2\psi(x, \xi_2)+\ldots+c_r\psi(x, \xi_r)+\ldots,$$

where $c_1, c_2, \ldots c_r, \ldots$ are numbers so chosen that the series $c_1+c_2+\ldots+c_r+\ldots$ is absolutely convergent.

Since the upper limits of all of the functions $|\psi(x, \xi)|$ have one and the same finite value, it follows that the series which defines $F(x)$ is uniformly convergent in the interval $(-\pi, \pi)$; and thus that the function $F(x)$ is continuous.

The expression

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} F(x')\, \frac{\sin\frac{1}{2}(2n+1)(x'-x)}{\sin\frac{1}{2}(x'-x)}\, dx'$$

is accordingly equal to the sum

$$\frac{1}{2\pi}\sum_1^\infty c_n \int_{-\pi}^{\pi} \psi(x', \xi_n)\frac{\sin\frac{1}{2}(2n+1)(x'-x)}{\sin\frac{1}{2}(x'-x)}\, dx,$$

which may be written in the form

$$c_1\chi_1(x, n) + c_2\chi_2(x, n) + \dots + c_r\chi_r(x, n) + \dots.$$

We have $\lim\limits_{n=\infty} \chi_1(x, n) = \psi(x, \xi_1)$, unless $x = \xi_1$; at which point the limit may be either 0 or ∞, according to the mode in which n is indefinitely increased, or may have other values between 0 and ∞. A similar statement holds as regards $\lim\limits_{n=\infty} \chi_2(x, n)$ at the point ξ_2; and generally $\lim\limits_{n=\infty} \chi_r(x, n)$ is $\psi_r(x, \xi_r)$, unless $x = \xi_r$, in which case the limit depends upon the mode in which n becomes infinite.

At the point ξ_r, the term $c_r\chi_r(x, n)$ has its limit indefinitely great, provided n is indefinitely increased in a proper manner; but it might happen that the limit of

$$c_{r+1}\chi_{r+1}(\xi_r, n) + c_{r+2}\chi_{r+2}(\xi_r, n) + \dots$$

is also infinite, although each term has a finite limit. In that case the limit of the whole expression for the sum of the series might be finite, or zero, in whatever manner n were made to become indefinitely great. If this happened for a particular set of values of $c_1, c_2, \dots c_r, \dots$, it would no longer happen if these numbers were replaced by

$$c_1e_1,\ c_2e_1e_2,\ c_3e_1e_2e_3, \dots,\ c_re_1e_2\dots e_r, \dots,$$

where $e_1, e_2, e_3, \dots$ is a properly chosen sequence of diminishing positive numbers. For, if

$$c_r\chi_r(\xi_r, n) + \{c_{r+1}\chi_{r+1}(\xi_r, n) + c_{r+2}\chi_{r+2}(\xi_r, n) + \dots\}$$

had a finite limit, when n is indefinitely increased, being dependent on the form $\infty - \infty$, the expression

$$e_1e_2\dots e_rc_r\chi_r(\xi_r, n) + e_1e_2\dots e_{r+1}\{c_{r+1}\chi_{r+1}(\xi_r, n) + c_{r+2}e_{r+2}\chi_{r+2}(\xi_r, n) + \dots\}$$

would also have a finite limit, or become zero, only in case

$$e_{r+1}\frac{c_{r+1}\chi_{r+1}(\xi_r, n) + c_{r+2}e_{r+2}\chi_{r+2}(\xi_r, n) + \dots}{c_{r+1}\chi_{r+1}(\xi_r, n) + c_{r+2}\chi_{r+2}(\xi_r, n) + \dots}$$

had unity as its limit, when n is indefinitely increased. But this limit can be altered by changing e_{r+1} without altering $e_{r+2}, e_{r+3}, \ldots$; and thus e_{r+1} can certainly be so chosen that this expression does not converge to unity, when n is indefinitely increased.

It has therefore been shewn that, by altering the numbers $c_1, c_2, c_3, \ldots$ in a suitable manner, the infinite limit of $c_r \chi_r(\xi_r, n)$ will no longer be removed by means of an infinite limit of the sum

$$c_{r+1}\chi_{r+1}(\xi_r, n) + c_{r+2}\chi_{r+2}(\xi_r, n) + \ldots.$$

It has thus been shewn that *it is possible so to choose the numbers* $c_1, c_2, c_3, \ldots$, *that the continuous function* $F(x) = \sum_1^\infty c_r \psi(x, \xi_r)$ *is such that its Fourier's series fails to converge at each point of the everywhere-dense set* $\{\xi_n\}$.

THE SERIES OF ARITHMETIC MEANS RELATED TO FOURIER'S SERIES.

468. It has recently been shewn* that a divergent series may be utilized in various ways for the representation of a function, and consequently for the calculation of approximate values of the function. The simplest of these methods is that due to Césaro, of taking the arithmetic means of the partial sums of the series.

If s_n be the nth partial sum of the convergent series of numbers

$$a_1 + a_2 + \ldots + a_n + \ldots,$$

and S_n denote the arithmetic mean $\frac{1}{n}(s_1 + s_2 + \ldots + s_n)$, then S_n converges to the sum of the given convergent series; or $\lim_{n=\infty} S_n = \lim_{n=\infty} s_n = s$. To prove this theorem, let ϵ be an arbitrarily chosen positive number, and let r be an integer such that $|s - s_{r+1}|, |s - s_{r+2}|, \ldots$ are all less than ϵ. We have now, if $n > r$,

$$S_n - s = (s_1 + s_2 + \ldots + s_r)\frac{1}{n} + \left\{\frac{s_{r+1} + s_{r+2} + \ldots + s_n}{n} - s\right\};$$

and keeping r fixed, n may be so chosen that $(s_1 + s_2 + \ldots + s_r)\frac{1}{n}$ is numerically less than ϵ; also the second term in the expression for $S_n - s$ is equal to $-\frac{r}{n}s + \eta$, where $|\eta| < \epsilon$, and by taking n sufficiently great, this is numerically less than 2ϵ. It follows that $|S_n - s| < 3\epsilon$, from and after some fixed value of n. Therefore, since ϵ is arbitrary, we have $\lim_{n=\infty} S_n = s$.

* For an account of these methods, see Borel's *Leçons sur les séries divergentes*.

It may however happen that, when the series Σa_n is not convergent, S_n still converges to a definite number z; this number z may then be regarded as the sum of the divergent series, in an extended sense of the term "sum." For example, in the case of the oscillating series $1-1+1-1+\ldots$, S_n converges to zero, which may therefore be regarded, in the new sense, as the sum of the series.

If, however, Σa_n diverges to ∞, or to $-\infty$, and is not oscillatory, S_n must also diverge. For $S_{n+m} > \dfrac{s_1+s_2+\ldots+s_n}{n+m} + \dfrac{m}{m+n}N$, if n can be so fixed that $s_{n+1}, s_{n+2}, \ldots$ all exceed the positive number N. Therefore $\lim\limits_{m=\infty} S_{n+m} \geqq N$; and since N may be taken arbitrarily great, if Σa_n diverges to $+\infty$, it follows that $\lim\limits_{n=\infty} S_n = +\infty$.

From the point of view of the theory of sets of points, it may happen that the points $P_1, P_2, \ldots P_n, \ldots$ which represent $s_1, s_2, \ldots s_n, \ldots$ do not converge to a single limiting point, but that the set of points $\bar{P}_1, \bar{P}_2, \ldots \bar{P}_n, \ldots$, where $\bar{P}_n$ is the centroid of the points $P_1, P_2, \ldots P_n$, has a single limiting point $\bar{P}$, which represents the number z.

In the case of a series $u_1(x)+u_2(x)+\ldots+u_n(x)+\ldots$ involving a variable x with a given domain, it may happen that the series fails to converge for some, or all, of the values of x, but that the function $S(x)$, of which the value for each value of x is the sum of the series, in the extended sense, has a definite value in the whole domain of x.

469. This method of summation has been applied by Fejér* to the case of Fourier's series. It can be shewn that, in a large class of cases, the Fourier's series corresponding to a function $f(x)$, when summed by the method of arithmetic means, converges to the value $\frac{1}{2}\lim\limits_{h=0}\{f(x+h)+f(x-h)\}$, when this expression has a definite meaning; no assumption being made as to the convergence of the Fourier's series when summed in the ordinary manner.

Since $s_n(x) = \dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi} f(x')\,dx' + \frac{1}{\pi}\sum_{s=1}^{n-1}\int_{-\pi}^{\pi} f(x')\cos s(x-x')\,dx'$,

we have $$S_n(x) = \frac{1}{\pi}\int_{-\pi}^{\pi}\left\{\frac{1}{2n} + \sum_{s=1}^{n-1}\frac{n-s}{n}\cos s(x-x')\right\} f(x')\,dx';$$

from which it is easily found that

$$S_n(x) = \frac{1}{2n\pi}\int_{-\pi}^{\pi}\left\{\frac{\sin\frac{1}{2}n(x'-x)}{\sin\frac{1}{2}(x'-x)}\right\}^2 f(x')\,dx'.$$

* *Math. Annalen*, vol. LVIII; also *Comptes Rendus*, for December 1900, and for April 1902. Fejér considered only the case in which the function has a Riemann integral. Lebesgue extended the result to the more general case; see the *Leçons sur les séries de Fourier*, p. 94.

Writing $x' = x + 2z$, and remembering that $f(x')$ is defined as a periodic function, for values of x' not necessarily in the interval $(-\pi, \pi)$, we have

$$S_n(x) = \frac{1}{n\pi}\int_0^{\frac{1}{2}\pi} \left(\frac{\sin nz}{\sin z}\right)^2 \{f(x+2z) + f(x-2z)\}\, dz.$$

It is easily seen that

$$\frac{1}{n\pi}\int_0^{\frac{1}{2}\pi} \left(\frac{\sin nz}{\sin z}\right)^2 dz = \frac{1}{2};$$

therefore we have

$$S_n(x) - \tfrac{1}{2}\lim_{h=0}\,[f(x+h) + f(x-h)] = \frac{1}{n\pi}\int_0^{\frac{1}{2}\pi} \left(\frac{\sin nz}{\sin z}\right)^2 F(z)\, dz,$$

where $F(z)$ denotes

$$f(x+2z) + f(x-2z) - \lim_{h=0}\{f(x+h) + f(x-h)\},$$

and it is assumed that x is a point at which $\lim\limits_{h=0}\{f(x+h)+f(x-h)\}$ has a definite finite value.

The expression $$\frac{1}{n\pi}\int_a^{\frac{1}{2}\pi} \left(\frac{\sin nz}{\sin z}\right)^2 F(z)\, dz$$

will be first examined, in order to find whether it tends to a definite limit, as n is indefinitely increased. The number α is fixed, and such that $0 < \alpha \leqq \frac{1}{2}\pi$.

We have

$$\frac{1}{n\pi}\int_a^{\frac{1}{2}\pi} \left(\frac{\sin nz}{\sin z}\right)^2 F(z)\, dz = \frac{1}{2n\pi}\int_a^{\frac{1}{2}\pi} \frac{1}{\sin^2 z} F(z)\, dz - \frac{1}{2n\pi}\int_a^{\frac{1}{2}\pi} \frac{\cos 2nz}{\sin^2 z} F(z)\, dz;$$

and $$\frac{1}{2n\pi}\int_a^{\frac{1}{2}\pi} \frac{1}{\sin^2 z} F(z)\, dz$$

is less, in absolute value, than

$$\frac{1}{2n\pi \sin^2 \alpha}\int_0^{\frac{1}{2}\pi} |F(z)|\, dz,$$

or than $$\frac{1}{2n\pi \sin^2\alpha}\left[\int_{-\pi}^{\pi} |f(x)|\, dx + \tfrac{1}{2}\pi\, \left|\lim_{h=0}\{f(x+h)+f(x-h)\}\right|\right].$$

It follows that, at any point x, at which $\lim\limits_{h=0}\{f(x+h)+f(x-h)\}$ is definite, n can be so chosen that

$$\left|\frac{1}{2n\pi}\int_a^{\frac{1}{2}\pi} \frac{1}{\sin^2 z} F(z)\, dz\right| < \epsilon.$$

It also follows that, in any interval of x, contained in another interval in which $f(x)$ is limited, and such that $\lim\limits_{h=0}\{f(x+h)+f(x-h)\}$ has everywhere a definite value, this inequality is satisfied for all values of x in the interval, provided n has a sufficiently great value, independent of x. Again since

$\operatorname{cosec}^2 z$ has limited total fluctuation in the interval $(\alpha, \frac{1}{2}\pi)$, it follows from the theorem in § 458, that

$$\int_\alpha^{\frac{1}{2}\pi} \frac{\cos 2nz}{\sin^2 z} f(x \pm 2z)\, dz$$

converges to zero, uniformly for all values of x. It is thus seen that

$$\frac{1}{2n\pi}\int_\alpha^{\frac{1}{2}\pi} \frac{\cos 2nz}{\sin^2 z} F(z)\, dz$$

converges to zero, at any point at which $\lim\limits_{h=0}\{f(x+h)+f(x-h)\}$ exists and is finite; and that the convergence is uniform in an interval contained in another interval in which $f(x)$ is limited, and in which the limit is everywhere definite. The following theorem has now been established:—

If $f(x)$ be either limited, or unlimited, but possess a Lebesgue integral in the interval $(-\pi, \pi)$, then $\frac{1}{n\pi}\int_\alpha^{\frac{1}{2}\pi}\left(\frac{\sin nz}{\sin z}\right)^2 F(z)\, dz$, where $0 < \alpha < \frac{1}{2}\pi$, converges to zero, as n is indefinitely increased, at any point x at which

$$\lim_{h=0}\{f(x+h)+f(x-h)\}$$

has a definite finite value. The convergence is uniform in any interval (a, b), in which $f(x)$ is limited, it being assumed that $\lim\limits_{h=0}\{f(x+h)+f(x-h)\}$ has definite values everywhere in (a, b), including the end-points a and b.

We now have to investigate the limiting value of

$$\frac{1}{n\pi}\int_0^{\frac{1}{2}\pi}\left(\frac{\sin nz}{\sin z}\right)^2 F(z)\, dz,$$

which may be expressed as

$$\frac{1}{n\pi}\int_0^{\pi/(2n+1)} F(z)\left(\frac{\sin nz}{\sin z}\right)^2 dz + \frac{1}{n\pi}\int_{\pi/(2n+1)}^{\alpha} F(z)\left(\frac{\sin nz}{\sin z}\right)^2 dz$$
$$+\frac{1}{n\pi}\int_\alpha^{\frac{1}{2}\pi} F(z)\left(\frac{\sin nz}{\sin z}\right)^2 dz,$$

where $\pi/(2n+1) < \alpha < \frac{1}{2}\pi$.

The first part of this expression is numerically less than

$$\frac{n}{\pi}\int_0^{\pi/(2n+1)} |F(z)|\, dz,$$

or than $\frac{n}{2n+1} \times$ the upper limit of $|F(z)|$ in the interval $\left(0, \frac{\pi}{2n+1}\right)$. At a point x at which $\lim\limits_{h=0}\{f(x+h)+f(x-h)\}$ is finite and definite, n may be chosen so great that the upper limit of $|F(z)|$ in $\left(0, \frac{\pi}{2n+1}\right)$ is less than the positive arbitrarily chosen number ϵ. Moreover, in virtue of the theorem established in § 185, in any interval (a, b), in which $f(x)$ is limited, and in which $\lim_{h=0} \{f(x+h) + f(x-h)\}$ has everywhere definite values, including the ends of the intervals, n may be so chosen that the upper limit of $|F(z)|$ in $\left(0, \frac{\pi}{2n+1}\right)$ is less than ϵ, for that value, and for all greater values, of n, and for all values of x in the interval. Therefore n_1 can be determined, such that

$$\left|\frac{1}{n\pi}\int_0^{\pi/(2n+1)} F(z)\left(\frac{\sin nz}{\sin z}\right)^2 dz\right| < \epsilon,$$

for $n \geqq n_1$, and for all values of x in an interval in which the specified conditions are satisfied.

The second part of the expression is numerically less than

$$\frac{1}{n\pi}\int_{\pi/(2n+1)}^{\alpha} \left(\frac{\pi}{2}\right)^2 \frac{1}{z^2} |F(z)|\, dz;$$

or than $\frac{\pi}{4n} \bar{F} \int_{\pi/(2n+1)}^{\alpha} \frac{1}{z^2} dz$, which is less than $\frac{3}{4}\bar{F}$; where $\bar{F}$ is the upper limit of $|F(z)|$ in the interval $(0, \alpha)$. Now, for any value of x, at which $\lim_{h=0} \{f(x+h) + f(x-h)\}$ has a finite value, α may be so chosen that $\frac{3}{4}\bar{F} < \epsilon$; also one and the same value of α will satisfy this condition, for all values of x in an interval in which $\lim_{h=0} \{f(x+h) + f(x-h)\}$ everywhere exists, including the end-points, and in which $f(x)$ is limited.

The number α having been so fixed, and n_1 then fixed as above, and such that $\frac{\pi}{2n_1+1} < \alpha$, n_2 can be determined so that

$$\left|\frac{1}{n\pi}\int_{\alpha}^{\frac{1}{2}\pi} F(z)\left(\frac{\sin nz}{\sin z}\right)^2 dz\right| < \epsilon,$$

for $n \geqq n_2$, and for all values of x in the specified interval.

If $\bar{n}$ be the greater of the two integers n_1, n_2, we now have

$$\left|\frac{1}{n\pi}\int_0^{\frac{1}{2}\pi} \left(\frac{\sin nz}{\sin z}\right)^2 F(z)\, dz\right| < 3\epsilon,$$

for $n \geqq \bar{n}$, at a point x at which $\lim_{h=0} \{f(x+h) + f(x-h)\}$ has a definite finite value. Moreover $\bar{n}$ will, if properly chosen, suffice for all points x, in an interval in which $f(x)$ is limited, and in which $\lim_{h=0} \{f(x+h) + f(x-h)\}$ has everywhere a definite finite value, including the end-points of the interval.

The following theorem has now been established:—

If $f(x)$ be any function, limited or unlimited in $(-\pi, \pi)$, which has a Lebesgue integral, and therefore a corresponding Fourier's series, the function

$S_n(x)$, which is the arithmetic mean of the first n terms of the Fourier's series, converges to $\frac{1}{2}\lim_{h=0}\{f(x+h)+f(x-h)\}$ at any point x, at which this limit has a definite finite value, as n is indefinitely increased. Moreover, the convergence is uniform in any interval (a, b) in which f(x) is limited, and in which $\lim_{h=0}\{f(x+h)+f(x-h)\}$ has everywhere, including a and b, definite finite values.

In particular, $S_n(x)$ converges to $f(x)$ at any point of continuity of $f(x)$; and it converges to $f(x)$ uniformly in any interval in which $f(x)$ is continuous, the continuity existing at the end-points, on both sides of those points.

No assumption has been made as regards the convergence of the Fourier's series when summed in accordance with the ordinary method.

470. Let the upper and the lower limits of indeterminacy of $s(x)$, for a Fourier's series at a point x, be denoted by $\underline{s(x)}$ and $\overline{s(x)}$; either of these may have one of the improper values $+\infty$, or $-\infty$. It is easily seen that the points $s_1(x), s_2(x), \ldots s_n(x)\ldots$ must be everywhere-dense in the interval $(\underline{s(x)}, \overline{s(x)})$; in fact this holds for any series for which the limit of the nth term is zero, as n is indefinitely increased. For, if ϵ be positive, and arbitrarily small, n may be so chosen that $|s_n(x)-s_{n+1}(x)|$, $|s_{n+1}(x)-s_{n+2}(x)|, \ldots$ are all $<\epsilon$; therefore in the interval $(\underline{s(x)}, \overline{s(x)})$, there exists no sub-interval of length ϵ which contains no points of $\{s_n(x)\}$. Since ϵ is arbitrarily small, it follows that the points $s_n(x)$ are everywhere-dense in the interval $(\underline{s(x)}, \overline{s(x)})$.

There can only be a finite number of values of n, for which $s_n(x)$ does not lie in the interval $(\underline{s(x)}-\epsilon, \overline{s(x)}+\epsilon)$; let this number of values be r, and let the sum of the corresponding values of $s_n(x)$ be Σ. We have then

$$S_n(x)-\overline{s(x)}=\frac{1}{n}\{s_1(x)+s_2(x)+\ldots+s_n(x)\}-\overline{s(x)},$$

$$<\frac{(n-r)\{\overline{s(x)}+\epsilon\}+\Sigma}{n}-\overline{s(x)}<\frac{\Sigma}{n}+\epsilon-\frac{r}{n}\{\overline{s(x)}+\epsilon\}<2\epsilon,$$

if n be taken sufficiently great. Therefore, from and after some fixed value of n, $S_n(x)$ is less than $\overline{s(x)}+2\epsilon$. Similarly, it can be shewn that $S_n(x)$ is greater than $\underline{s(x)}-2\epsilon$, from and after some fixed value of n. Since ϵ is arbitrary, it follows that, when $S_n(x)$ converges to a definite number $S(x)$, as n is indefinitely increased, $S(x)$ lies in the interval $(\underline{s(x)}, \overline{s(x)})$.

From this result the following theorem now follows:—

If $f(x)$ be summable in $(-\pi, \pi)$, then, at any point x at which $\lim_{h=0}\{f(x+h)+f(x-h)\}$ has a definite finite value, the upper and lower limits *of indeterminacy of the sum of the Fourier's series, corresponding to* $f(x)$, *form a limited, or an unlimited, interval which contains the number* $\frac{1}{2}\lim_{h=0}\{f(x+h)+f(x-h)\}$ *to which the arithmetic mean* $S_n(x)$ *of the first* n *partial sums of the series converges. In particular, at a point of continuity of the function,* $f(x)$ *is in the interval of which the ends are the limits of indeterminacy of the series at the point* x.

It follows, as a particular case of this theorem, that, *at a point* x *at which the Fourier's series converges, and at which* $\frac{1}{2}\lim_{h=0}\{f(x+h)+f(x-h)\}$ *exists as a finite number, the Fourier's series converges to that number. In particular, at a point of continuity of* $f(x)$, *if the series converges, then it must converge to the value* $f(x)$.

It has been shewn above that, for a point x, the numbers $s_1(x), s_2(x), \ldots s_n(x) \ldots$ are everywhere-dense in the limited, or unlimited, interval $(\underline{s(x)}, \overline{s(x)})$, of indeterminacy of the sum of the series. It follows that, if ξ be any chosen point in this interval, a sequence $s_{n_1}(x), s_{n_2}(x), \ldots s_{n_r}(x) \ldots$ can be determined, which converges to the number ξ. At a point x, at which $\{S_n(x)\}$ converges, we may take ξ to coincide with the value to which the sequence converges; and we thus obtain the following theorem :—

At a point x, *at which* $\lim_{h=0}\{f(x+h)+f(x-h)\}$ *has a definite finite value, the Fourier's series corresponding to* $f(x)$ *can be replaced, by bracketing the terms of the series in a suitable manner, and amalgamating the terms in each bracket, by a convergent series of which the sum is* $\frac{1}{2}\lim_{h=0}\{f(x+h)+f(x-h)\}$. *In particular, at a point of continuity of* $f(x)$, *a suitable sequence of partial sums of the Fourier's series converges to the value* $f(x)$.

In the particular case of a function $f(x)$ which has either a Riemann integral, or an absolutely convergent Harnack integral, in $(-\pi, \pi)$, the set of points at which $\lim_{h=0}\{f(x+h)+f(x-h)\}$ has not a definite value, forms a set of points of measure zero. It therefore follows* that, *for such a function, the set of points at which it is impossible, by bracketing the terms of the series suitably, to convert the series into a convergent series of which the sum is* $\frac{1}{2}\lim_{h=0}\{f(x+h)+f(x-h)\}$, *has the measure zero.* The requisite system of bracketing is in general dependent upon the particular point.

471. The method of § 469 can be applied to the case in which

* See Hobson, *Proc. Lond. Math. Soc.*, ser. 2, vol. III, p. 55, where this is established for the case of a limited function, by a different method.

$\lim_{h=0}\{f(x+h)+f(x-h)\}$ has no definite value at the point x. Let $F_1(z)$ denote $f(x+2z)+f(x-2z)$; we then have

$$\frac{1}{n\pi}\int_0^{\frac{1}{2}\pi}\left(\frac{\sin nz}{\sin z}\right)^2 F_1(z)\,dz$$

$$=\frac{1}{n\pi}\int_0^{\alpha} F_1(z)\left(\frac{\sin nz}{\sin z}\right)^2 dz+\frac{1}{n\pi}\int_{\alpha}^{\frac{1}{2}\pi} F_1(z)\left(\frac{\sin nz}{\sin z}\right)^2 dz.$$

Now α can be so chosen that $F_1(z)-\overline{F_1}(+0)<\epsilon$, and $F_1(z)-\underline{F_1}(+0)>\epsilon$, for all values of z such that $0<z\leqq\alpha$. We then have

$$\frac{1}{n\pi}\int_0^{\frac{1}{2}\pi}\left(\frac{\sin nz}{\sin z}\right)^2 F_1(z)\,dz<\frac{1}{n\pi}\{\epsilon+\overline{F_1}(+0)\}\int_0^{\alpha}\left(\frac{\sin nz}{\sin z}\right)^2 dz+\frac{1}{n\pi}\int_{\alpha}^{\frac{1}{2}\pi}\left(\frac{\sin nz}{\sin z}\right)^2 dz.$$

In accordance with the first result in § 469, the second term on the right-hand side converges to zero, as n is indefinitely increased. Also the first term converges to $\frac{1}{2}\{\epsilon+\overline{F_1}(+0)\}$, since $\frac{1}{n\pi}\int_0^{\frac{1}{2}\pi}\left(\frac{\sin nz}{\sin z}\right)^2 dz=\frac{1}{2}\pi$. Hence we have $\overline{\lim}_{n=\infty} S_n(x)\leqq\frac{1}{2}\{\epsilon+\overline{F_1}(+0)\}$, and, since ϵ is arbitrarily small, we have

$\overline{\lim}_{n=\infty} S_n(x)\leqq\frac{1}{2}\overline{F_1}(+0)$; and in a similar manner, $\underline{\lim}_{n=\infty} S_n(x)\geqq\frac{1}{2}\underline{F_1}(+0)$.

It therefore appears that $\overline{\lim}_{n=\infty} S_n(x)$, $\underline{\lim}_{n=\infty} S_n(x)$, both lie in the interval $(\frac{1}{2}\underline{F_1}(+0), \frac{1}{2}\overline{F_1}(+0))$, which is certainly in the interval of which the ends are

$$\tfrac{1}{2}\{\underline{f(x+0)}+\underline{f(x-0)}\},\ \tfrac{1}{2}\{\overline{f(x+0)}+\overline{f(x-0)}\}.$$

If $S_n(x)$ converge at the point x, the value to which it converges must lie in the interval of which the ends are these two numbers. If the Fourier's series converge at the point x, it necessarily converges to the same value to which $S_n(x)$ converges.

We have therefore the following theorem:—

If a Fourier's series be convergent at a point at which $f(x)$ has a discontinuity of the second kind, its sum at the point lies between $\frac{1}{2}\{\overline{f(x+0)}+\overline{f(x-0)}\}$ *and* $\frac{1}{2}\{\underline{f(x+0)}+\underline{f(x-0)}\}$, *or may be equal to one of these numbers.*

If, at a point x, the Fourier's series diverges to either $+\infty$, or to $-\infty$, but does not oscillate, then $S_n(x)$ diverges to the same improper limit. The above discussion shews that this can only happen when $F_1(z)$ has its upper limit, or its lower limit, for $z=0$, indefinitely great.

It therefore follows that *the Fourier's series can only diverge to $+\infty$, or to $-\infty$ and be non-oscillatory, at a point at which the function has an infinite discontinuity.*

The series may oscillate between infinite limits of indeterminacy at a point at which the functional limits are all finite.

The results here obtained include the theorem of § 469, as the special case which arises when $\overline{F_1}(+0)=\underline{F_1}(+0)$.

PROPERTIES OF THE COEFFICIENTS OF FOURIER'S SERIES.

472. The coefficients

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,dx,\quad a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,dx,\quad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,dx,$$

which occur in the Fourier's series which corresponds to a function $f(x)$, may be termed the *Fourier's constants* related to the function $f(x)$. They possess important properties which may be regarded as connected with a general theory of the Fourier's constants related to a given function which possesses a Lebesgue integral. These properties are independent of any assumptions as to the convergence of the series; and thus the relation* of the constants to the function may be denoted by

$$f(x) \sim \tfrac{1}{2}a_0 + (a_1\cos x + b_1\sin x) + \ldots + (a_n\cos nx + b_n\sin nx) + \ldots.$$

The following important property of the Fourier's constants will be here established, for the case of a limited function which is integrable† in accordance with Riemann's definition:—

‡*The function $f(x)$ being limited, and integrable in accordance with Riemann's definition, in the interval $(-\pi, \pi)$, the series $\tfrac{1}{2}a_0^2 + \sum\limits_{s=1}^{\infty}(a_s^2 + b_s^2)$ converges to the value $\dfrac{1}{\pi}\displaystyle\int_{-\pi}^{\pi}\{f(x)\}^2\,dx$; where a_s, b_s denote the Fourier's constants corresponding to $f(x)$.*

The function $S_n(x)$ denoting, as in § 469, the arithmetic mean of the first n terms of the series, and $S(x)$ denoting $\lim\limits_{n=\infty} S_n(x)$, it has been seen in § 469, that $S(x) = f(x)$, at a point of continuity of $f(x)$. The points of discontinuity of $f(x)$ form a set of points of measure zero. At one of these points, $S(x)$ has a definite value if $\lim\limits_{h=0}\{f(x+h) + f(x-h)\}$ exists at the point; otherwise $S(x)$ may be indeterminate between finite limits of indeterminacy, the upper one of which does not exceed $\tfrac{1}{2}\{\overline{f(x+0)} + \overline{f(x-0)}\}$, and the lower one of which is not less than $\tfrac{1}{2}\{\underline{f(x+0)} + \underline{f(x-0)}\}$. The function $S(x)$ is therefore limited, and determinate, except at points of a set of zero measure. It follows§, since

$$\{f(x) - S(x)\}^2 = \lim_{n=\infty}\{f(x) - S_n(x)\}^2,$$

* See Hurwitz, *Math. Annalen*, vol. LVII, 1903, p. 426, where this terminology was introduced for the case of a function possessing a Riemann integral.

† The theorem has been extended by Lebesgue to the case of functions which are integrable in accordance with his definition. See the *Leçons sur les séries trigonométriques*, pp. 100, 101.

‡ See de la Vallée-Poussin, *Annales de la soc. scien. de Bruxelles*, vol. XVII, p. 18. Also Hurwitz, *loc. cit.*

§ This involves an extension of Lebesgue's theorem given in § 384, that, if $s(x) = \lim\limits_{n=\infty} s_n(x)$, where $s_n(x)$ is limited for all values of n and x, then $\int_a^b s(x)\,dx = \lim\limits_{n=\infty}\int_a^b s_n(x)\,dx$, to the case in that $$\lim_{n=\infty}\int_{-\pi}^{\pi}\{f(x)-S_n(x)\}^2dx=\int_{-\pi}^{\pi}\{f(x)-S(x)\}^2dx=0.$$

Since $$S_n(x)=\tfrac{1}{2}a_0+\sum_{s=1}^{s=n}\frac{n-s}{n}(a_s\cos sx+b_s\sin sx);$$
we find that

$$\int_{-\pi}^{\pi}\{f(x)-S_n(x)\}^2dx=\int_{-\pi}^{\pi}\{f(x)\}^2dx-\pi\left\{\tfrac{1}{2}a_0^2+\sum_{s=1}^{n}\frac{n^2-s^2}{n^2}(a_s^2+b_s^2)\right\}.$$

We then have also,

$$\begin{aligned}\int_{-\pi}^{\pi}&\{f(x)-S_n(x)\}^2dx\\&=\int_{-\pi}^{\pi}\{f(x)\}^2dx-\pi\left\{\tfrac{1}{2}a_0^2+\sum_{s=1}^{s=n}(a_s^2+b_s^2)\right\}+\frac{\pi}{n^2}\sum_{s=1}^{s=n}s^2(a_s^2+b_s^2)\\&=\int_{-\pi}^{\pi}\left[f(x)-\tfrac{1}{2}a_0-\sum_{s=1}^{s=n}(a_s\cos sx+b_s\sin sx)\right]^2dx+\frac{\pi}{n^2}\sum_{s=1}^{n}s^2(a_s^2+b_s^2).\end{aligned}$$

Since $\lim_{n=\infty}\int_{-\pi}^{\pi}\{f(x)-S_n(x)\}^2dx=0$, we see, from the last expression, that $\lim_{n=\infty}\frac{\pi}{n^2}\sum_{s=1}^{n}s^2(a_s^2+b_s^2)=0$; and therefore, from the first expression, it follows that

$$\int_{-\pi}^{\pi}\{f(x)\}^2dx-\pi\left\{\tfrac{1}{2}a_0^2+\sum_{s=1}^{s=n}(a_s^2+b_s^2)\right\}$$

converges to zero, as n is indefinitely increased. Therefore the series $\tfrac{1}{2}a_0^2+\sum_{s=1}(a_s^2+b_s^2)$ converges to the value $\frac{1}{\pi}\int_{-\pi}^{\pi}\{f(x)\}^2dx$.

From this theorem, Hurwitz has deduced the following more general theorem which was first discovered by Liapounoff* and by de la Vallée-Poussin:—

If $f(x)$, $\phi(x)$ be two limited integrable functions, then

$$\tfrac{1}{2}a_0a_0'+\sum_{s=1}^{\infty}(a_sa_s'+b_sb_s')$$

which, although $s(x)$ is limited, it may be indeterminate at points in (a, b) belonging to a set E of measure zero. To prove this extension, let $\Sigma_n(x)$ be a function which $=s_n(x)$ at every point of $C(E)$, and $=\overline{s(x)}$ at each point of E. Then $\overline{s(x)}=\lim_{n=\infty}\Sigma_n(x)$, everywhere in (a, b). Since $\Sigma_n(x)-s_n(x)$ is zero, except at the points of a set of zero measure, it follows that $\Sigma_n(x)$ is summable, since $s_n(x)$ is so. Therefore $\int_a^b\overline{s(x)}\,dx=\lim_{n=\infty}\int_a^b\Sigma_n(x)\,dx=\lim_{n=\infty}\int_a^b s_n(x)\,dx$; and since $\int_a^b s(x)\,dx=\int_a^b\overline{s(x)}\,dx$, it follows that $\int_a^b s(x)\,dx=\lim_{n=\infty}\int_a^b s_n(x)\,dx$.

* Stekloff states in the *Comptes Rendus* for Nov. 10, 1902, that the theorem was communicated by Liapounoff to the Kharkow Mathematical Society in 1896. The theorem was also obtained by de la Vallée-Poussin, *Annales de la soc. sci. de Bruxelles*, vol. XVII, p. 18.

converges absolutely to the value $\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,\phi(x)\,dx$; *where* a_0, a_s, b_s *are the Fourier's constants for* $f(x)$, *and* a_0', a_s', b_s' *those for* $\phi(x)$.

That the theorem holds in the case in which the Fourier's series are uniformly convergent is shewn by a direct formation of the product, and has long been known.

To establish the result, we observe that the two series

$$\tfrac{1}{2}(a_0+a_0')^2+\sum_{s=1}^{\infty}\{(a_s+a_s')^2+(b_s+b_s')^2\},$$

$$\tfrac{1}{2}(a_0-a_0')^2+\sum_{s=1}^{\infty}\{(a_s-a_s')^2+(b_s-b_s')^2\},$$

are absolutely convergent, and have for their sums

$$\frac{1}{\pi}\int_{-\pi}^{\pi}\{f(x)+\phi(x)\}^2\,dx,\quad \frac{1}{\pi}\int_{-\pi}^{\pi}\{f(x)-\phi(x)\}^2\,dx,$$

respectively. By taking the difference of these two series, the result follows at once.

This result may be applied to express the Fourier's constants of $f(x)\,\phi(x)$, the product of two limited integrable functions, in terms of the Fourier's constants for the two functions $f(x)$, $\phi(x)$.

Thus, if

$$f(x)\sim\tfrac{1}{2}a_0+\sum_{n=1}^{\infty}(a_n\cos nx+b_n\sin nx),$$

$$\phi(x)\sim\tfrac{1}{2}a_0'+\sum_{n=1}^{\infty}(a_n'\cos nx+b_n'\sin nx),$$

$$f(x)\,\phi(x)\sim\tfrac{1}{2}\alpha_0+\sum_{n=1}^{\infty}(\alpha_n\cos nx+\beta_n\sin nx),$$

we have, from the above theorem,

$$\alpha_0=\tfrac{1}{2}a_0a_0'+\sum_{r=1}^{\infty}(a_ra_r'+b_rb_r').$$

Corresponding to the function $\phi(x)\cos mx$, the Fourier's constant

$$\frac{1}{\pi}\int_{-\pi}^{\pi}\phi(x)\cos mx\cos rx\,dx=\tfrac{1}{2}(a'_{m+r}+a'_{m\sim r});$$

also the Fourier's constant for the function $f(x)\,\phi(x)\cos mx$, which corresponds to α_0 relative to $f(x)\,\phi(x)$, is α_m. We thus obtain an expression for α_m, by employing the two functions $f(x)$, $\phi(x)\cos mx$, and applying the above theorem. We then find that

$$\alpha_m=\tfrac{1}{2}a_0a_m'+\sum_{r=1}^{\infty}\{\tfrac{1}{2}a_r(a'_{m+r}+a'_{m\sim r})+\tfrac{1}{2}b_r(b'_{m+r}+b'_{m\sim r})\}.$$

In a similar manner, it can be proved that

$$\beta_m=\tfrac{1}{2}a_0b_m'+\sum_{r=1}^{\infty}\{\tfrac{1}{2}a_r(b'_{m+r}-b'_{m\sim r})-\tfrac{1}{2}b_r(a'_{m+r}-a'_{m\sim r})\}.$$

THE INTEGRATION OF FOURIER'S SERIES.

473. If $f(x)$ be any summable function which has a Fourier's series

$$\tfrac{1}{2}a_0 + \sum_{r=1}^{\infty} (a_r \cos rx + b_r \sin rx),$$

the integral $\int_{-\pi}^{x} f(x)\,dx$ is a continuous function, with limited total fluctuation in $(-\pi, \pi)$. It is therefore representable by a Fourier's series

$$\tfrac{1}{2}a_0' + \sum_{r=1}^{\infty} (a_r' \cos rx + b_r' \sin rx),$$

which is uniformly convergent in any interval in the interior of the interval $(-\pi, \pi)$. Denoting the function $\int_{-\pi}^{x} f(x)\,dx$ by $\phi(x)$, we have

$$\frac{1}{\pi}\int_{-\pi}^{\pi} \phi(x)\cos rx\,dx = \frac{1}{\pi}\left[\phi(x)\frac{\sin rx}{r}\right]_{-\pi}^{\pi} - \frac{1}{r\pi}\int_{-\pi}^{\pi} f(x)\sin rx\,dx,$$

or

$$a_r' = -\frac{1}{r}b_r,$$

whenever the function $f(x)$ is such that the formula of integration by parts is applicable. This has been shewn in § 394, to be the case when $f(x)$ is either a limited summable function, or also when it is integrable, but has points of infinite discontinuity which belong to a reducible set of points.

In a similar manner, it can be shewn that $b_r' = \frac{1}{r}a_r - \frac{1}{r}a_0 \cos r\pi$. Therefore the function $\phi(x) = \int_{-\pi}^{x} f(x)\,dx$ is represented by the Fourier's series

$$\tfrac{1}{2}a_0' + \sum_{r=1}^{\infty} \frac{1}{r}\left[-b_r \cos rx + (a_r - a_0 \cos r\pi)\sin rx\right].$$

To determine a_0', we observe that, at the point $x = -\pi$, the sum of the series must be $\frac{1}{2}\{\phi(-\pi+0) + \phi(\pi-0)\}$, or $\frac{1}{2}\pi a_0$; therefore

$$\tfrac{1}{2}a_0' - \sum_{r=1}^{\infty} \frac{b_r}{r}\cos r\pi = \tfrac{1}{2}\pi a_0.$$

Also, in the interior of the interval $(-\pi, \pi)$, we have

$$-\tfrac{1}{2}x = \sum_{r=1}^{\infty} \frac{1}{r}\cos r\pi \sin rx.$$

Therefore $\phi(x)$ is represented, in the interval $(-\pi, \pi)$, by

$$\tfrac{1}{2}a_0(\pi + x) + \Sigma \frac{1}{r}\{a_r \sin rx + b_r(\cos r\pi - \cos rx)\},$$

which is obtained by integrating the terms of the Fourier's series corresponding to $f(x)$, between the limits $-\pi$, x.

The following theorem has been established:—

If $f(x)$ be a summable function, which, if unlimited, has points of infinite discontinuity belonging only to a reducible set of points, then $\int_\alpha^\beta f(x)\,dx$, where $-\pi \leqq \alpha < \beta \leqq \pi$, is represented by the convergent series obtained by integration, term by term, of the Fourier's series corresponding to $f(x)$.

It will be observed that no assumptions have been made as regards the convergence of the Fourier's series which represents $f(x)$.

It appears from the theorem of § 472, that, when $f(x)$ is a limited function, integrable in accordance with Riemann's definition, the two series $\sum_{r=1}^{\infty} a_r^2$, $\sum_{r=1}^{\infty} b_r^2$ are both convergent. From this result*, the convergence of the two series $\sum_{r=1}^{\infty} \frac{1}{r}|a_r|$, $\sum_{r=1}^{\infty} \frac{1}{r}|b_r|$ follows. For $a_r^2 + \frac{1}{r^2} \geqq \frac{2}{r}|a_r|$, hence

$$\sum_1^r \frac{1}{r}|a_r| < \sum_1^r a_r^2 + \sum_1^r \frac{1}{r^2};$$

and thus $\sum_1^r \frac{1}{r}|a_r|$ is, for all values of r, less than some fixed positive number; whence the convergence of the series $\Sigma \frac{1}{r}|a_r|$ follows.

PROPERTIES OF POISSON'S INTEGRAL.

474. Let the function $f(x)$, defined for the interval $(-\pi, \pi)$, be either limited, or unlimited, but such that it possesses a Lebesgue integral in the interval, which integral is of course absolutely convergent. It has been pointed out in § 435, that Poisson's integral

$$\int_{-\pi}^{\pi} \frac{1-h^2}{1-2h\cos(x-x')+h^2} f(x')\,dx',$$

where $-1 < h < 1$, represents the sum of the convergent series

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x')\,dx' + \sum_{n=1}^{\infty} h^n \left\{\cos nx \,.\, \frac{1}{\pi}\int_{-\pi}^{\pi} \cos nx' \,.\, f(x')\,dx' \right.$$
$$\left. + \sin nx \,.\, \frac{1}{\pi}\int_{-\pi}^{\pi} \sin nx' \,.\, f(x')\,dx' \right\}.$$

It will here be shewn† that, as h converges to the limit 1, Poisson's integral converges to the value

$$\lim_{t=0} \tfrac{1}{2}\{f(x+t)+f(x-t)\},$$

* See Bôcher's "Theory of Fourier's series," *Annals of Math.*, ser. 2, vol. VII, p. 108.

† The limit to which Poisson's integral converges has been studied by Schwarz, in two memoirs; see his *Math. Abh.*, vol. II, pp. 144, 175. Schwarz has considered the case, more general than that in the text, in which x varies as well as h; but he has confined his attention to the case in which the function is either continuous, or else has only a finite number of discontinuities. See also Forsyth's *Theory of Functions*, 2nd ed. p. 450, or Picard's *Traité d'Analyse*, vol. I, p. 249. For a more complete treatment of questions connected with Poisson's integral, see the memoir by Fatou in the *Acta Mat.*, vol. XXX.

where x has any constant value such that this latter limit has a definite value.

We find, by a direct method, that

$$\int_0^\theta \frac{1-h^2}{1-2h\cos\theta+h^2}\,d\theta = 2\tan^{-1}\left(\frac{1+h}{1-h}\tan\tfrac{1}{2}\theta\right);$$

and thence that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1-h^2}{1-2h\cos\theta+h^2}\,d\theta = 1.$$

Denoting the value of Poisson's integral, for a fixed value of x, by $I(x)$, we have

$$I(x) - \lim_{t=0} \tfrac{1}{2}\{f(x+t)+f(x-t)\}$$
$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1-h^2}{1-2h\cos(x-x')+h^2}\,\phi(x')\,dx';$$

where $\phi(x')$ denotes

$$f(x') - \lim_{t=0} \tfrac{1}{2}\{f(x+t)+f(x-t)\},$$

and x has a fixed value such that

$$\lim_{t=0} \tfrac{1}{2}\{f(x+t)+f(x-t)\}$$

has a definite value.

A positive number δ can be so chosen that, if $0 < \xi \leqq \delta$,

$$|f(x+\xi)+f(x-\xi) - \lim_{t=0}\{f(x+t)+f(x-t)\}| < \epsilon,$$

where ϵ is a prescribed positive number. We have then

$$\left|\frac{1}{2\pi}\int_{x-\delta}^{x+\delta} \frac{1-h^2}{1-2h\cos(x-x')+h^2}\,\phi(x')\,dx'\right| < \frac{\epsilon}{2\pi}\int_0^\delta \frac{1-h^2}{1-2h\cos\xi+h^2}\,d\xi;$$

and the expression on the right-hand side is less than

$$\frac{\epsilon}{2\pi}\int_{-\pi}^{\pi} \frac{1-h^2}{1-2h\cos(x-x')+h^2}\,dx', \text{ or than } \epsilon.$$

The remaining part of the integral which represents

$$I(x) - \lim_{t=0} \tfrac{1}{2}\{f(x+t)+f(x-t)\}$$

is
$$\frac{1}{2\pi}\left[\int_{x+\delta}^{\pi} + \int_{-\pi}^{x-\delta}\right] \frac{1-h^2}{1-2h\cos(x-x')+h^2}\,\phi(x')\,dx';$$

and this is numerically less than

$$\frac{1-h^2}{1-2h\cos\delta+h^2}\cdot\frac{1}{2\pi}\int_{-\pi}^{\pi} |\phi(x')|\,dx',$$

or than

$$\frac{1-h^2}{1-2h\cos\delta+h^2}\left[\left|\lim_{t=0}\tfrac{1}{2}\{f(x+t)+f(x-t)\}\right|+\frac{1}{2\pi}\int_{-\pi}^{\pi}|f(x')|\,dx'\right].$$

Also $$\frac{1-h^2}{1-2h\cos\delta+h^2}<\frac{1-h}{(1-h)^2+4h\sin^2\frac{1}{2}\delta}<\frac{1-h}{4h\sin^2\frac{1}{2}\delta};$$

if then η be a prescribed positive number, we have

$$\frac{1-h^2}{1-2h\cos\delta+h^2}<\eta,\ \text{if}\ \frac{1-h}{4h\sin^2\frac{1}{2}\delta}<\eta,\ \text{or if}\ h>(1+4\eta\sin^2\tfrac{1}{2}\delta)^{-1},$$

and this is satisfied if

$$h>1-4\eta\sin^2\tfrac{1}{2}\delta.$$

The numbers ϵ, η being arbitrarily fixed, δ can be determined as above; then, provided $1-h<4\eta\sin^2\frac{1}{2}\delta$, we have

$$\left|I(x)-\tfrac{1}{2}\lim_{t=0}\{f(x+t)+f(x-t)\}\right|<\epsilon+A\eta,$$

where A is a fixed positive number, for a fixed value of x. Since ϵ, η are arbitrarily small, we have therefore proved that

$$\lim_{h=1}I(x)=\lim_{t=0}\tfrac{1}{2}\{f(x+t)+f(x-t)\}.$$

For, if ζ be arbitrarily fixed, we may first choose ϵ, so that $\epsilon<\frac{1}{2}\zeta$, and we may then choose $\eta<\zeta/2A$; hence, if h be sufficiently small, $I(x)$ differs from the limit by less than ζ.

The following theorem has now been established:—

If $f(x)$ be a limited or unlimited function, possessing a Lebesgue integral in the interval $(-\pi,\pi)$, then, for any fixed value of x, for which

$$\lim_{t=0}\tfrac{1}{2}\{f(x+t)+f(x-t)\}$$

has a definite value, Poisson's integral converges to the value of that limit, as h converges to 1. In particular, at a point of ordinary discontinuity of $f(x)$, Poisson's integral converges to the value

$$\tfrac{1}{2}\{f(x+0)+f(x-0)\};$$

and, at a point of continuity of $f(x)$, it converges to the value $f(x)$.

It has been already pointed out in § 435, that no conclusion can be drawn as to the convergence or non-convergence of the Fourier's series at such a point x. In case however the Fourier's series converges, it follows from Abel's theorem (§ 356), that it must converge to the same limit as does Poisson's integral. We therefore obtain the following theorem:—

If $f(x)$ be any function, for which the Fourier's coefficients exist, as proper (Lebesgue, or Riemann) integrals, or as absolutely convergent (Lebesgue, or

Harnack) improper integrals, then at any point x at which the Fourier's series is convergent, it must converge to the value

$$\lim_{t=0} \tfrac{1}{2}\{f(x+t)+f(x-t)\},$$

provided this limit have a definite value.

This theorem has already been established otherwise, in § 470.

It may be remarked that, if (α, β) be any interval in which $f(x)$ is continuous, the end-points α, β being points of continuity, then the number δ, corresponding to a fixed ϵ, may be chosen so as to be independent of x, for all values of x in (α, β). This follows from the property of uniform continuity of a continuous function (§ 175).

Also A is less than a fixed number, for all values of x in (α, β). It therefore follows that *Poisson's integral converges to the value $f(x)$ uniformly in the interval (α, β), in which $f(x)$ is continuous.*

APPROXIMATE REPRESENTATION OF FUNCTIONS BY FINITE TRIGONOMETRICAL SERIES.

475. If the function $f(x)$, defined for the interval $(-\pi, \pi)$, be continuous in the interval (α, β), contained in $(-\pi, \pi)$, including the end-points α, β, it has been seen in § 474, that Poisson's integral converges to the value $f(x)$, uniformly in the interval (α, β), as h converges to the value 1.

Therefore, a value h_1, of h, may be chosen, corresponding to an arbitrarily fixed positive number ϵ, so that $f(x)$ differs from the sum of the convergent series

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x')\,dx' + \sum_{n=1}^{\infty} h_1^n \left\{\cos nx \,.\, \frac{1}{\pi}\int_{-\pi}^{\pi} f(x')\cos nx'\,dx' + \sin nx \,.\, \frac{1}{\pi}\int_{-\pi}^{\pi} f(x')\sin nx'\,dx'\right\}$$

by less than $\frac{1}{2}\epsilon$, for all values of x in (α, β). Since the series converges uniformly for all values of x, an integer m may be so fixed, that the remainder of the series after the mth term is numerically less than $\frac{1}{2}\epsilon$, for all values of x. In this manner, we obtain* *a finite trigonometrical series*

$$\tfrac{1}{2}A_0 + (A_1\cos x + B_1 \sin x) + \ldots + (A_m \cos mx + B_m \sin mx)$$

the sum of which differs from $f(x)$ by less than ϵ, for every value of x in the interval (α, β) in which $f(x)$ is continuous.

This mode of approximate representation of $f(x)$, in the interval (α, β), is clearly not unique, because the values of the function in that part of $(-\pi, \pi)$

* See Picard's *Traité d'Analyse*, 2nd ed., vol. I, p. 275.

which is not in (α, β) may be altered in any manner, subject only to the integrability of $f(x)$ in $(-\pi, \pi)$, and the continuity of $f(x)$ at the points α, β.

In the above finite series, each of the circular functions can be expanded in powers of x, and the result rearranged as a power-series, of which the sum consequently differs from $f(x)$ by less than ϵ, for all values of x in (α, β). Since the power-series is uniformly convergent, we thus obtain a proof of Weierstrass' theorem, already established in § 373, that a finite polynomial $P(x)$ can be determined, such that $|f(x) - P(x)| < 2\epsilon$, for all values of x in (α, β); the number ϵ being arbitrarily chosen.

Another method*, not involving the use of Poisson's integral, may be employed to determine an approximate representation of a function $f(x)$, continuous in (α, β), by means of finite trigonometrical series. Choose l, so that $-l < \alpha < \beta < l$. As in § 373, a continuous polygonal line can be constructed, such that its ordinate, for each point x in (α, β), differs from $f(x)$ by less than $\frac{1}{2}\epsilon$. The polygonal line may be extended to the whole interval $(-l, l)$, so as to be a continuous polygonal line for the whole interval, and to be such that its ordinates at the points $x = l, -l$ are equal to one another. In virtue of Dirichlet's theory of Fourier's series, the polygonal line may be represented, for the whole interval $(-l, l)$, by a Fourier's series

$$\tfrac{1}{2}a_0 + \sum_{n=1}^{\infty}\left(a_n \cos\frac{n\pi x}{l} + b_n \sin\frac{n\pi x}{l}\right);$$

and, by the theorem of § 451, this series converges uniformly in $(-l, l)$, to the value $f(x)$. The sum of the Fourier's series differs from $f(x)$ by less than $\frac{1}{2}\epsilon$, at every point of (α, β). The integer m may be so chosen that the sum of the terms for $n > m$, is less than $\frac{1}{2}\epsilon$, for all values of x in (α, β), on account of the uniform convergence. Therefore the finite series

$$\tfrac{1}{2}a_0 + \sum_{n=1}^{n=m}\left(a_n \cos\frac{n\pi x}{l} + b_n \sin\frac{n\pi x}{l}\right),$$

has the required property, that its sum differs from $f(x)$ by less than ϵ, for all values of x in (α, β). This method may be applied, in the same manner as in the case of the preceding one, to prove Weierstrass' theorem relating to the approximate representation of a continuous function by a finite polynomial.

476. Let $f(x)$ be a function such that both $f(x)$ and $\{f(x)\}^2$ possess Lebesgue integrals in the interval $(-\pi, \pi)$; and let $s_m(x)$ denote the sum of a finite trigonometrical series

$$\tfrac{1}{2}A_0 + \sum_{n=1}^{n=m}(A_n \cos mx + B_n \sin mx).$$

* Volterra, *Rendiconti del Circolo mat. di Palermo*, vol. XI, 1897, p. 83.

Let us consider the integral

$$I_m = \int_{-\pi}^{\pi} \{f(x) - s_m(x)\}^2 dx.$$

We find that

$$I_m = \int_{-\pi}^{\pi} \{f(x)\}^2 dx + \pi \left[\frac{1}{2}\left\{A_0 - \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\, dx\right\}^2 + \sum_{n=1}^{n=m}\left\{A_n - \frac{1}{\pi}\int_{-\pi}^{\pi} f(x) \cos nx dx\right\}^2 + \sum_{n=1}^{n=m}\left\{B_n - \frac{1}{\pi}\int_{-\pi}^{\pi} f(x) \sin nx dx\right\}^2\right] - \frac{1}{2\pi}\left\{\int_{-\pi}^{\pi} f(x)\, dx\right\}^2 - \frac{1}{\pi}\sum_{n=1}^{n=m}\left[\left\{\int_{-\pi}^{\pi} f(x) \cos nx dx\right\}^2 + \left\{\int_{-\pi}^{\pi} f(x) \sin nx dx\right\}^2\right].$$

If I_m be regarded as a quadratic function of

$$A_0,\ A_1,\ B_1 \ldots A_m,\ B_m,$$

it is clear that the value of I_m will be least, when

$$A_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\, dx, \quad A_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x) \cos nx dx,$$

$$B_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x) \sin nx dx,$$

for $n = 1, 2, 3, \ldots m$; *i.e.* when A_0, A_n, B_n are the Fourier's coefficients corresponding to the function $f(x)$. These values of A_0, A_n, B_n are therefore such that the finite trigonometrical series gives the best approximation to the value of $f(x)$, in accordance with the standard of the method of least squares. The following theorem has been now established:—

If $f(x)$ be defined for the interval $(-\pi, \pi)$, and be such that both the function itself, and its square, possess Lebesgue integrals in the interval, then the values of the $2m + 1$ constants $A_0, A_1, B_1 \ldots A_m, B_m$, which are such that*

$$\int_{-\pi}^{\pi} \left[f(x) - \tfrac{1}{2}A_0 - \sum_{n=1}^{n=m} (A_m \cos mx + B_m \sin mx)\right]^2 dx,$$

has the smallest value, are the Fourier's coefficients corresponding to the function $f(x)$.

The minimum value of the integral I_m is

$$\int_{-\pi}^{\pi} \{f(x)\}^2 dx - \pi\left[\tfrac{1}{2}a_0^2 + \sum_{n=1}^{n=m} (a_n^2 + b_n^2)\right],$$

where a_0, a_n, b_n denote the Fourier's constants corresponding to the function $f(x)$. It follows that this difference is essentially positive, whatever value m may have, and therefore the series $\frac{1}{2}a_0^2 + \sum_{n=1}^{\infty} (a_n^2 + b_n^2)$ is necessarily con-

* This theorem was given by Toepler, in a somewhat less general form, see *Wiener Anzeigen*, vol. XIII, 1876.

vergent. It has been shewn in § 472, that, on certain assumptions, the series converges to the value $\frac{1}{\pi}\int_{-\pi}^{\pi}\{f(x)\}^2 dx$. An attempt was made by Harnack* to establish this fact directly, and to found thereon a theory of the convergence of Fourier's series.

It follows, from the above result, that the series $\sum_1^{\infty} a_n^2$, $\sum_1^{n} b_n^2$ are both convergent, and therefore that $\lim_{n=\infty} a_n = 0$, $\lim_{n=\infty} b_n = 0$, which has already been established in § 454, independently of the assumption here made, that $\{f(x)\}^2$ is integrable in $(-\pi, \pi)$.

THE DIFFERENTIATION OF FOURIER'S SERIES.

477. In general, the series obtained by differentiating a convergent Fourier's series is not convergent, as may, for example, be seen in the case of the series $\sum_1^{\infty}\frac{1}{n}\sin nx$; neither is the series so obtained necessarily the Fourier's series corresponding to $f'(x)$.

Let $f(x)$ be a limited function, with only a finite number of ordinary discontinuities; let it also be assumed that $f'(x)$ has a Lebesgue integral in $(-\pi, \pi)$, and that, if it have points of infinite discontinuity, such points form a reducible set. This is consistent with there being a set of points of zero measure at which $f'(x)$ has no definite value. At the points of discontinuity of $f(x)$, we may regard $f'(x)$ as undefined. We have then

$$\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,dx = \left[\frac{1}{n\pi}f(x)\sin nx\right]_{-\pi}^{\pi} - \frac{1}{n\pi}\int_{-\pi}^{\pi} f'(x)\sin nx\,dx$$

$$= \frac{1}{n\pi}\left[-\Sigma\{f(\alpha+0)-f(\alpha-0)\}\sin n\alpha\right] - \frac{1}{n\pi}\int_{-\pi}^{\pi} f'(x)\sin nx\,dx,$$

the summation Σ referring to the finite number of points α of ordinary discontinuity of $f(x)$ in the interior of $(-\pi, \pi)$. In a similar manner, we find that

$$\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,dx$$

$$= \frac{1}{n\pi}\left[(-1)^n\{f(-\pi+0)-f(\pi-0)\} + \Sigma\{f(\alpha+0)-f(\alpha-0)\}\cos n\alpha\right]$$

$$+\frac{1}{n\pi}\int_{-\pi}^{\pi} f'(x)\cos nx\,dx.$$

Also

$$\frac{1}{\pi}\int_{-\pi}^{\pi} f'(x)\,dx = \frac{1}{\pi}\left[f(\pi-0)-f(-\pi+0)-\Sigma\{f(\alpha+0)-f(\alpha-0)\}\right].$$

* See two articles in the *Math. Annalen*, vol. XVII.

If then, the Fourier's coefficients for the functions $f(x)$, $f'(x)$ be denoted by a_0, a_n, b_n, and a_0', a_n', b_n' respectively, we have

$$a_0' = \frac{1}{\pi}[f(\pi - 0) - f(-\pi + 0)] - \frac{1}{\pi}\Sigma\{f(\alpha + 0) - f(\alpha - 0)\},$$

$$a_n' = nb_n - \frac{1}{\pi}[(-1)^n\{f(-\pi + 0) - f(\pi - 0)\} + \Sigma\{f(\alpha + 0) - f(\alpha - 0)\}\cos n\alpha],$$

$$b_n' = -na_n - \frac{1}{\pi}\Sigma\{f(\alpha + 0) - f(\alpha - 0)\}\sin n\alpha.$$

In particular, if $f(x)$ be continuous in the interval $(-\pi, \pi)$, so that the function obtained by extending $f(x)$ beyond the interval, in accordance with the rule $f(x) = f(x \pm 2\pi)$, is continuous except at the points $-\pi, \pi$, we have

$$a_0' = \frac{1}{\pi}\{f(\pi) - f(-\pi)\}, \quad a_n' = nb_n + \frac{(-1)^n}{\pi}\{f(\pi) - f(-\pi)\},$$

$b_n' = -na_n$. Unless $f(\pi) = f(-\pi)$, the Fourier's series corresponding to $f'(x)$ is not obtained by term by term differentiation of the Fourier's series for $f(x)$. Even when this condition is satisfied, no assertion can in general be made as to the convergence of the Fourier's series for $f'(x)$. We have thus obtained the following theorem:—

If $f(x)$ be continuous in $(-\pi, \pi)$, and if $f(-\pi) = f(\pi)$, and $f'(x)$ have a Lebesgue integral, and have at most a reducible set of points of infinite discontinuity, the Fourier's series for $f'(x)$, whether it converge or not, is obtained by the term by term differentiation of that corresponding to $f(x)$.

If it be known that $f'(x)$ has limited derivatives at any point, or if

$$\lim_{h=+0}\frac{f'(x+h) - f'(x+0)}{h}, \quad \lim_{h=+0}\frac{f'(x-h) - f'(x-0)}{-h}$$

are definite, or are indeterminate between finite limits of indeterminacy, then, in accordance with Theorem III, of § 457, the Fourier's series for $f'(x)$ converges at the point x.

478. In case the function $f(x)$ have derivatives $f'(x)$, $f''(x)$, ... of any number of orders, and $f(x)$, $f'(x)$, $f''(x)$, ... are all limited, and continuous in $(-\pi, \pi)$, except at a finite number of points at which they have ordinary discontinuities, the coefficients a_n, b_n may be expressed in a form which exhibits these discontinuities.

At a point α of discontinuity of $f(x)$, the function $f'(x)$ may be regarded as undefined, the values of $f'(\alpha + 0)$, $f'(\alpha - 0)$ being

$$\lim_{h=+0}\frac{f(\alpha+h) - f(\alpha+0)}{h}, \quad \lim_{h=+0}\frac{f(\alpha-h) - f(\alpha-0)}{-h}$$

respectively. A similar remark applies to the higher differential coefficients.

We find, by integrating twice by parts,

$$a_n = -\frac{1}{n\pi}\Sigma\{f(\alpha+0)-f(\alpha-0)\}\sin n\alpha - \frac{1}{n^2\pi}\Sigma\{f'(\beta+0)-f'(\beta-0)\}\cos n\beta - \frac{1}{n^2\pi}\int_{-\pi}^{\pi} f''(x)\cos nx\,dx,$$

$$b_n = \frac{1}{n\pi}\Sigma\{f(\alpha+0)-f(\alpha-0)\}\cos n\alpha - \frac{1}{n^2\pi}\Sigma\{f'(\beta+0)-f'(\beta-0)\}\sin n\beta - \frac{1}{n^2\pi}\int_{-\pi}^{\pi} f''(x)\sin nx\,dx,$$

where $-\pi$ is now included among the points α of discontinuity of $f(x)$, and amongst β, the points of discontinuity of $f'(x)$. The points α in general occur amongst the points β.

We may proceed, by further* integration by parts, to express a_n and b_n in a series proceeding by powers of $1/n$, the coefficients of which involve the measures of discontinuity of the functions at the points $\alpha, \beta, \ldots$.

Conversely, if the Fourier's coefficients for $f(x)$ are given in the forms

$$a_n = \frac{1}{n}\Sigma A\sin n\alpha + \frac{1}{n^2}\Sigma B\cos n\beta + \ldots,$$

$$b_n = -\frac{1}{n}\Sigma A\cos n\alpha + \frac{1}{n^2}\Sigma B\sin n\beta - \ldots,$$

so that the Fourier's series has for its general term

$$\frac{1}{n}\Sigma A\sin n(\alpha - x) + \frac{1}{n^2}\Sigma B\cos n(\beta - x) + \ldots,$$

we have

$$f(\alpha+0)-f(\alpha-0) = -\pi A,\ f'(\beta+0)-f'(\beta-0) = -\pi B, \ldots.$$

Thus the points of discontinuity, and the measures of discontinuity of $f(x)$, $f'(x)$, ... are determined when a_n, b_n are exhibited as series proceeding according to powers of $1/n$.

479. The following further theorems† relating to the differentiation of trigonometrical series will be stated:—

If the trigonometrical series

$$\tfrac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n\cos nx + b_n\sin nx)$$

converge for a particular value c of x, and if the series

$$\sum_{n=1}^{\infty}(-na_n\sin nx + nb_n\cos nx),$$

* See Stokes "On the critical values of the sums of periodic series," *Math. and Phys. Papers*, vol. I, where this investigation is carried out in detail, and the resulting formulae for the differentiation of Fourier's series are applied to physical problems.

† See Bôcher's "Introduction to the theory of Fourier's series," *Annals of Math.*, vol. VII, p. 120. The second theorem is substantially due to Lerch, *Annales sc. de l'école normale*, ser. 3, vol. XI, p. 351.

obtained by term by term differentiation, converge uniformly in an interval (α, β) *which contains the point c in its interior, then the original series converges uniformly in* (α, β), *and the function* $f(x)$ *represented by it has, throughout the interval, a differential coefficient represented by the derived series.*

If the series $$\sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$$

converge for a particular value c of x, which is not zero or a multiple of π, *and if* $\lim_{n=\infty} a_n = 0$, $\lim_{n=\infty} b_n = 0$, *then throughout an interval* (α, β) *which contains the point c in its interior, but does not include the point* 0, *or* $k\pi$, *where k is any integer, the series converges uniformly in the interval* (α, β), *and the function* $f(x)$ *represented by it will have a differential coefficient* $f'(x)$ *given by*

$$2 \sin x \,.\, f'(x) = \sum_{n=0}^{\infty} \{[(n-1)\, a_{n-1} - (n+1)\, a_{n+1}] \cos nx + [(n-1)\, b_{n-1} - (n+1)\, b_{n+1}] \sin nx\},$$

where $a_{-1} = b_{-1} = a_0 = b_0 = 0$, *provided this last series converges uniformly in the interval* (α, β).

It is clear that, for a function $f(x)$ which possesses differential coefficients of all orders in the interval $(-\pi, \pi)$, it is not in general possible to obtain representations of all these differential coefficients by means of successive term by term differentiation of the Fourier's series which represents $f(x)$. The following theorem, due to Borel*, gives the means of obtaining the requisite representation of such functions:—

Having given a function $f(x)$ *which has differential coefficients of all orders throughout the interval* $(-\pi, \pi)$, *the function can be represented by means of a series of the type*

$$\sum_{n=0}^{\infty} (A_n x^n + a_n \cos nx + b_n \sin nx)\,;$$

and the differential coefficients of $f(x)$, *of all orders, are represented by the series obtained by successive term by term differentiation of this series. All the series so obtained converge uniformly in the interval* $(-\pi, \pi)$.

GENERAL EXAMPLES.

1. The trigonometrical series

$$b_1 \sin x + b_2 \sin 2x + \ldots + b_n \sin nx + \ldots,$$

is uniformly convergent in any interval not containing the point $x=0$, or any point $x = \pm 2k\pi$, (k integral), if $\lim_{n=\infty} b_n = 0$, and if also $\sum_{n=1}^{\infty} |b_n - b_{n+1}|$ be convergent. For

$$2 \sin \tfrac{1}{2}x \,.\, s_n(x) = b_1 \cos \tfrac{1}{2}x - \sum_{r=1}^{n-1} (b_r - b_{r+1}) \cos \tfrac{1}{2}(2r+1)\, x - b_n \cos \tfrac{1}{2}(2n+1)\, x,$$

* See the *Leçons sur les fonctions de variables réelles*, p. 68, where this theorem is proved.

whence the result follows. The series* converges for all values of x, if $\lim_{n=\infty} b_n = 0$, and if also $b_n \geqq b_{n+1}$, for all values of n greater than some fixed value m; the convergence is then uniform in any interval which does not contain $x=0$ or $x = \pm 2k\pi$, for any integral value of k.

The series $\frac{1}{2}a_0 + a_1 \cos x + a_2 \cos 2x + \ldots$, may similarly be shewn to converge uniformly in any interval not containing $x=0$, or any point $\pm 2k\pi$, if $\lim_{n=\infty} a_n = 0$, and if also $\sum_{n=1}^{\infty} |a_n - a_{n+1}|$ be convergent. If $\lim_{n=\infty} a_n = 0$, and $a_n \geqq a_{n+1}$, for $n > m$, the series* converges for all values of x, except 0 or $\pm 2k\pi$.

2. Let $f(x)$† be a function, of period 2π, limited and integrable in any interval which does not contain the point $x=0$, or any point $x = 2k\pi$; but let $f(x)$ not satisfy these conditions in the neighbourhood of $x=0$. Let it be assumed, (1) that $|f(x) + f(-x)|$ is integrable, in $(0, \pi)$, (2) that $\lim_{x=0} \{xf(x)\} = 0$, and, (3) that $xf(x)$ has its Fourier's series convergent at the point $x=0$. The Fourier's coefficients a_n, b_n, for the function $f(x)$, then exist, and $\lim_{n=\infty} a_n = 0$. Also it follows from (3), that $\lim_{n=\infty} b_n = 0$. For this last condition is equivalent to

$$\lim_{n=\infty} \int_{-\pi}^{\pi} f(x') \tan \tfrac{1}{2}x' \frac{\sin \frac{1}{2}(2n+1)x'}{\sin \frac{1}{2}x'} dx' = 0,$$

which holds if $f(x) \tan \frac{1}{2}x$ have its Fourier's series convergent at $x=0$; and $f(x) \tan \frac{1}{2}x$ may clearly be replaced by $xf(x)$.

It can now be seen easily that

$$\int_{-\pi}^{\pi} \frac{\sin \frac{1}{2}(2n+1)(x-x')}{\sin \frac{1}{2}(x-x')} f(x')\, dx'$$

has the limit 0, when n is indefinitely increased, on condition that, in the neighbourhood of $x'=0$, those elements which correspond to values of x' of opposite sign, but of equal values, are made to coalesce. When the conditions (1), (2), (3) are satisfied, the necessary and sufficient condition that the Fourier's series should converge to $f(x)$ is that that function which $= f(x)$ in the neighbourhood of the point x, and is elsewhere zero, should be representable by a Fourier's series.

Let
$$f(x) = \frac{\sin \dfrac{\pi}{x}}{x \log \dfrac{1}{x} \log \log \dfrac{1}{x}}, \text{ where } 0 < x \leqq a < e^{-e},$$

and let $f(x) + f(-x) = 0$. This function satisfies conditions (1), (2), (3), and is representable by a series

$$a_1 \sin \frac{\pi x}{a} + a_2 \sin \frac{2\pi x}{a} + \ldots.$$

$|f(x)|$ is not integrable, although $f(x)$ is so; thus the series is a generalized Fourier's series.

3. The convergent series† $\sum_{n=2}^{\infty} \frac{\sin nx}{\log n}$ represents a function which is not integrable, in any sense, in an interval containing the point $x=0$. The series $\sum_{n=2}^{\infty} \frac{-\cos nx}{n \log n}$ is not convergent.

* Schlömilch, *Compendium d. höheren Analysis*, vol. I, § 40.

† See Fatou, *Comptes rendus*, March 26, 1906.

4. In the series $\sum\limits_1^\infty \sin (n\,!\,\pi x)$, the coefficients do not become indefinitely small, and therefore the series is not a Fourier's series. The series converges, however, for all rational values of x; it also converges for an infinite number of irrational values, for example, for $x=\sin 1$, $\cos 1$, $2/e$, and for multiples of these values; also for odd multiples of e. This example is due to Riemann, and the series has been considered in detail by Genocchi*.

5. Consider the series $\sum\limits_{n=0}^\infty c_n \cos n^2 x$, $\sum\limits_{n=1}^\infty c_n \sin n^2 x$, where $c_0, c_1, c_2, \ldots$ are positive numbers, and such that $\lim\limits_{n=\infty} c_n = 0$, but such that $\sum\limits_{n=0}^\infty c_n$ is divergent. The points of convergence, and the points of divergence, of these series both form everywhere-dense sets. These series have been treated in detail by Genocchi.

6. The function $f(x) = \sum\limits_{n=1}^\infty \frac{1}{n}(nx)$, where (nx) denotes the excess of nx over the nearest integer, and where $(nx)=0$ when nx is half an odd integer, is not integrable in accordance with Riemann's definition. Riemann has however given the series

$$\frac{1}{\pi} \sum_{n=1}^\infty \frac{S_\theta\left[-(-1)^\theta\right]}{n} \sin 2n\pi x,$$

as representing $f(x)$; where the summation S_θ refers to all the factors θ, of n.

RIEMANN'S THEORY OF TRIGONOMETRICAL SERIES.

480. After the fundamental investigation of Dirichlet, in which sufficient conditions were obtained for the convergence of the Fourier's series corresponding to a given function, the next great advance in the theory was made by Riemann†, in his celebrated memoir on the representation of a function by means of trigonometrical series. This memoir formed the point of departure, on which much of the subsequent development of the theory depended.

Denoting such a series by

$$A_0 + A_1 + A_2 + \ldots + A_n + \ldots,$$

where
$$A_0 = \tfrac{1}{2}a_0, \quad A_n = a_n \cos nx + b_n \sin nx,$$

it is assumed, for the most part, that

$$\lim_{n=\infty} (a_n \cos nx + b_n \sin nx) = 0,$$

for each value of x in a given interval. It was proved later by Cantor, that this assumption necessarily implies that $\lim\limits_{n=\infty} a_n = 0$, and $\lim\limits_{n=\infty} b_n = 0$. In some

* *Atti di Torino*, vol. x.

† "Ueber die Darstellbarkeit einer Function durch eine trigonometrische Reihe." This memoir, originally written as a thesis in 1854, was published in the *Abhandlungen d. K. Ges. d. Wissenschf. zu Göttingen*, vol. XIII. See also Riemann's *Gesammelte Werke*, 2nd ed., p. 227.

parts of Riemann's investigations, it is sufficient to make only the wider assumption that

$$|a_n \cos nx + b_n \sin nx|$$

is less than some fixed positive number K, for all values of n, and for all values of x in a prescribed interval. It is not assumed that the coefficients necessarily have the form of the Fourier's coefficients; so that the theory refers to trigonometrical series in general.

For each value of x in the interval $(-\pi, \pi)$, for which the series converges, the limiting sum will be denoted by $f(x)$. This function $f(x)$, defined by means of the given series, is defined only for those values of x for which the series converges. In later investigations undertaken by Du Bois Reymond, it is assumed that the function $f(x)$ exists also at a point x at which the series oscillates; the function being regarded as multiple-valued at such a point, with limits of indeterminacy identical with those of the series at the point.

The question asked and answered by Riemann was as follows:—The function $f(x)$ being defined at the points of convergence of the given series, as the limiting sum of that series, what can be inferred as to the properties of the function $f(x)$?

In order to answer this question, Riemann undertook an examination of the properties of the function $F(x)$, which is represented by the series

$$C + C'x + \frac{1}{2}A_0x^2 - A_1 - \frac{1}{2^2}A_2 - \dots - \frac{1}{n^2}A_n - \dots,$$

obtained by integrating the terms of the given series twice.

For any value of x for which $|A_n|$ is less than some fixed positive number, for all values of n, the function $F(x)$ exists.

In any interval of x in which $|A_n|$ is less than a fixed positive number ϵ, for all values of x and n, the terms of the series are less, in absolute value, than those of

$$\epsilon\left(1 + \frac{1}{1^2} + \frac{1}{3^2} + \dots + \frac{1}{n^2} + \dots\right);$$

and therefore the series is uniformly convergent. It follows that the function $F(x)$ is continuous in the interval. This is, in particular, the case if

$$\lim_{n=\infty} a_n = 0, \ \lim_{n=\infty} b_n = 0.$$

The function $F(x)$ has the properties formulated in the following three theorems:—

(1) *For any value of x, for which the series $A_0 + A_1 + A_2 + \dots + A_n + \dots$ converges, the expression*

$$\frac{F(x+\alpha+\beta) - F(x+\alpha-\beta) - F(x-\alpha+\beta) + F(x-\alpha-\beta)}{4\alpha\beta}$$

converges to the value $f(x)$, when α, β are indefinitely diminished in any manner such that the ratio of one to the other remains finite.

In particular

$$\frac{F(x+2\alpha)+F(x-2\alpha)-2F(x)}{4\alpha^2}$$

converges to $f(x)$, as α is indefinitely diminished. In this theorem, it is sufficient to assume that, in a fixed neighbourhood of x

$$| a_n \cos nx + b_n \sin nx |$$

is less than a fixed positive number ϵ, for all values of n, and for all points in the neighbourhood.

(2) *For any value of x whatever*

$$\frac{F(x+2\alpha)+F(x-2\alpha)-2F(x)}{2\alpha}$$

converges to the limit zero, as α converges to zero.

It is unnecessary that the function $f(x)$ should exist for the value of x concerned, and it is sufficient that, for all values of x,

$$a_n \cos nx + b_n \sin nx$$

should have the limit zero, as n is indefinitely increased.

(3) *If b, c are two arbitrary constants, such that $b < c$, and if $\lambda(x)$, $\lambda'(x)$ are functions which are continuous in the interval (b, c), and vanish for $x = b$, $x = c$, and if further $\lambda(x)$, be such that $\lambda''(x)$ is* a limited and integrable function in the interval (b, c), then the expression*

$$\mu^2 \int_b^c F(x) \cos \mu (x-a) \,.\, \lambda(x)\, dx$$

converges to zero, as μ is indefinitely diminished, uniformly for all values of a. It is necessary that a_n and b_n have the limit zero.

481. Proceeding to the demonstration of these theorems, it may be observed that Riemann's theorem (1) can be generalised so as to include the case in which the series $A_0 + A_1 + A_2 + \ldots$ does not converge at the point x, but oscillates, with finite upper and lower limits U, L, of indeterminacy. In that case it will be shewn that the expression

$$\frac{F(x+2\alpha)-2F(x)+F(x-2\alpha)}{4\alpha^2}$$

has a limit, for $\alpha = 0$, which, whether it be definite or not†, lies between the numbers

$$\tfrac{1}{2}(U+L) \pm \tfrac{1}{2}(U-L)\left(1+\frac{1}{\pi}+\frac{1}{\pi^2}\right).$$

* Riemann restricts $\lambda''(x)$ to have only a finite number of maxima and minima in the interval, and makes no mention of uniform convergence.

† Du Bois Reymond gave the value $\frac{1}{2}(U+L) \pm \left(\frac{3}{2}+\frac{1}{\pi^2}+\frac{1}{\pi}\right)(U-L)$; see *Abh. d. bayerisch. Akad.* vol. XII, p. 136.

In the case $U=L$, this reduces to Riemann's original theorem. To prove the generalised theorem, we find that

$$\frac{F(x+2\alpha)-2F(x)+F(x-2\alpha)}{4\alpha^2}=A_0+A_1\left(\frac{\sin\alpha}{\alpha}\right)+A_2\left(\frac{\sin 2\alpha}{2\alpha}\right)^2+\dots.$$

Now $A_0+A_1+\dots+A_{n-1}$ lies between the two numbers

$$\tfrac{1}{2}(U+L)\pm\tfrac{1}{2}(U-L)\pm\epsilon_n;$$

where ϵ_n is a positive number, such that, if δ be any arbitrarily small prescribed positive number, a number m can be found, such that for $n \geqq m$, $\epsilon_n<\delta$. We may accordingly write

$$A_0+A_1+\dots+A_{n-1}=\tfrac{1}{2}(U+L)+\tfrac{1}{2}(U-L)\,\theta_n+\eta_n$$

where $1\geqq\theta_n\geqq-1$, and η_n is a number such that $|\eta_n|<\delta$, for $n\geqq m$.

The series $\sum\limits_{n=0}^{\infty}A_n\left(\frac{\sin n\alpha}{n\alpha}\right)^2$ may be written

$$\tfrac{1}{2}(U+L)+\sum_{n=1}^{\infty}\left(\frac{U-L}{2}\theta_n+\eta_n\right)\left\{\left(\frac{\sin\overline{n-1}\alpha}{\overline{n-1}\alpha}\right)^2-\left(\frac{\sin n\alpha}{n\alpha}\right)^2\right\};$$

and we may divide the series into three parts; first, from $n=1$ to $n=m$, where $|\eta_n|<\delta$, for $n\geqq m$; next, from $n=m+1$ to $n=s$, where s is the integral part of π/α; and lastly from $n=s+1$ onwards; we suppose α to be chosen so small that $m\alpha<\pi$. The first part of the sum consists of a number m, of terms, where m is independent of α; and this part has the limit zero, when α is indefinitely diminished. The second part of the sum is such that, in every term,

$$\left(\frac{\sin\overline{n-1}\alpha}{\overline{n-1}\alpha}\right)^2-\left(\frac{\sin n\alpha}{n\alpha}\right)^2$$

is positive; and therefore this part of the sum lies between the two values

$$\pm\left(\frac{U-L}{2}+\delta\right)\left\{\left(\frac{\sin m\alpha}{m\alpha}\right)^2-\left(\frac{\sin s\alpha}{s\alpha}\right)^2\right\}.$$

The third part of the sum may be written in the form

$$\sum_{n=s+1}^{\infty}\left(\frac{U-L}{2}\theta_n+\eta_n\right)\left\{\left(\frac{\sin\overline{n-1}\alpha}{\overline{n-1}\alpha}\right)^2-\left(\frac{\sin\overline{n-1}\alpha}{n\alpha}\right)^2\right\}$$
$$-\sum_{n=s+1}^{\infty}\left(\frac{U-L}{2}\theta_n+\eta_n\right)\frac{\sin(2n-1)\alpha\sin\alpha}{n^2\alpha^2}.$$

Now
$$\left(\frac{\sin\overline{n-1}\alpha}{\overline{n-1}\alpha}\right)^2-\left(\frac{\sin\overline{n-1}\alpha}{n\alpha}\right)^2$$

is positive, and less than

$$\left\{\frac{1}{(n-1)^2}-\frac{1}{n^2}\right\}\frac{1}{\alpha^2};$$

and $\frac{\sin\alpha\sin(2n-1)\alpha}{n^2\alpha^2}$ is numerically less than $\frac{1}{n^2\alpha}$; hence the sum of the third part of the series lies between

$$\left(\frac{U-L}{2}+\delta\right)\frac{1}{s^2\alpha^2}+\left(\frac{U-L}{2}+\delta\right)\frac{1}{s\alpha},$$

and the same expression with the sign changed. When α is indefinitely diminished, $s\alpha$ has the limit π; hence the limiting sum of the series

$$\Sigma A_n\left(\frac{\sin n\alpha}{n\alpha}\right)^2$$

lies between the two numbers

$$\tfrac{1}{2}(U+L)\pm\left(\frac{U-L}{2}+\delta\right)\left(1+\frac{1}{\pi}+\frac{1}{\pi^2}\right);$$

or, since δ can be chosen as small as we please, it lies between

$$\tfrac{1}{2}(U+L)\pm\tfrac{1}{2}(U-L)\left(1+\frac{1}{\pi}+\frac{1}{\pi^2}\right).$$

In case x is a point of convergence of the series ΣA_n, we have $U=L=f(x)$, and the above limit becomes $f(x)$.

To prove Riemann's original theorem, we have

$$F(x+\alpha+\beta)-2F(x)+F(x-\alpha-\beta)=(\alpha+\beta)^2[f(x)+\delta_1]$$
$$F(x+\alpha-\beta)-2F(x)+F(x-\alpha+\beta)=(\alpha-\beta)^2[f(x)+\delta_2]$$

where δ_1, δ_2 converge to zero, when $\alpha+\beta$, $\alpha-\beta$ do so. From these equations we find

$$\frac{F(x+\alpha+\beta)-F(x+\alpha-\beta)-F(x-\alpha+\beta)+F(x)}{4\alpha\beta}$$
$$=f(x)+\frac{(\alpha+\beta)^2}{4\alpha\beta}\delta_1-\frac{(\alpha-\beta)^2}{4\alpha\beta}\delta_2;$$

the expression on the right hand side converges to $f(x)$, when α, β are indefinitely diminished so that their ratio remains finite, since δ_1, δ_2 converge to zero; and therefore Riemann's original theorem is established.

482. To prove Riemann's second theorem that, whether ΣA_n converges or not, so long as $\lim_{n=\infty} A_n$, for each fixed x, is zero,

$$\frac{F(x+2\alpha)-2F(x)+F(x-2\alpha)}{2\alpha}$$

converges to zero, as α does so, we observe that the series $\sum_{n=0}^{\infty} A_n\left(\frac{\sin n\alpha}{n\alpha}\right)^2$ can be separated into three groups of terms. The first group is from $n=0$ to $n=m$, where m is so chosen that $A_n<\epsilon$, for $n>m$; the sum of these terms remains finite, as α is indefinitely diminished; denote this sum by S_1.

The second group of terms is from $n=m+1$ to $n=s$, where $s\alpha \leqq$ a fixed number $c<(s+1)\alpha$; the sum of these terms is numerically less than $\epsilon\frac{c}{\alpha}$. The remainder of the series is numerically

$$<\frac{\epsilon}{\alpha^2}\sum_{n=s+1}^{\infty}\frac{1}{n^2}<\frac{\epsilon}{s\alpha^2},$$

and $\frac{\epsilon}{s\alpha^2}$ differs arbitrarily little from $\frac{\epsilon}{\alpha c}$, when α is sufficiently small. It follows that

$$\left|\frac{F(x+2\alpha)-2F(x)+F(x-2\alpha)}{2\alpha}\right|=\left|2\alpha\Sigma A_n\left(\frac{\sin n\alpha}{n\alpha}\right)^2\right|$$
$$<2\left(|S_1|\alpha+\epsilon c+\frac{\epsilon}{c}\right),$$

whence the theorem follows, since ϵ is arbitrarily small.

There now remains Riemann's third theorem, which requires the consideration of the expression

$$\mu^2\int_b^c F(x)\cos\mu(x-a)\,\lambda(x)\,dx,$$

when μ becomes indefinitely great, the function $\lambda(x)$ satisfying the conditions already stated. Since $F(x)$ is represented by a uniformly convergent series, we may replace the integral by the sum of the integrals obtained by substituting, in the integral, the series which represents $F(x)$. We have then

$$\mu^2\int_b^c F(x)\cos\mu(x-a)\,\lambda(x)\,dx=\mu^2\int_b^c (C+C'x+\tfrac{1}{2}A_0x^2)\cos\mu(x-a)\,\lambda(x)\,dx$$
$$-\sum_{n=1}^{\infty}\frac{\mu^2}{n^2}\int_b^c A_n\cos\mu(x-a)\,\lambda(x)\,dx.$$

We shall first consider the series

$$\sum_{n=1}^{\infty}\frac{\mu^2}{n^2}\int_b^c A_n\cos\mu(x-a)\,\lambda(x)\,dx.$$

The general term of this series is

$$\frac{\mu^2}{2n^2}a_n\int_b^c[\cos\{(\mu+n)x-\mu a\}+\cos\{(\mu-n)x-\mu a\}]\,\lambda(x)\,dx$$
$$+\frac{\mu^2}{2n^2}b_n\int_b^c[\sin\{(\mu+n)x-\mu a\}-\sin\{(\mu-n)x-\mu a\}]\,\lambda(x)\,dx,$$

and each of these integrals may be integrated twice by parts, since

$$\lambda'(x),\quad \lambda''(x)$$

are both limited integrable functions. The general term then takes the form

$$-\frac{\mu^2}{2n^2(\mu+n)^2}a_n\int_b^c \cos\{(\mu+n)x-\mu a\}\lambda''(x)\,dx$$
$$-\frac{\mu^2}{2n^2(\mu-n)^2}a_n\int_b^c \cos\{(\mu-n)x-\mu a\}\lambda''(x)\,dx$$
$$-\frac{\mu^2}{2n^2(\mu+n)^2}b_n\int_b^c \sin\{(\mu+n)x-\mu a\}\lambda''(x)\,dx$$
$$+\frac{\mu^2}{2n^2(\mu-n)^2}b_n\int_b^c \sin\{(\mu-n)x-\mu a\}\lambda''(x)\,dx.$$

The series can be divided into four parts, containing respectively the four sets of terms here expressed; and these four parts will be considered separately. Corresponding to an arbitrarily chosen positive number ϵ, a value μ_1, of μ can be so chosen that

$$\int_b^c \cos(\mu+n)x\,.\,\lambda''(x)\,dx,\quad \int_b^c \sin(\mu+n)x\,.\,\lambda''(x)\,dx$$

are both numerically $<\frac{1}{2}\epsilon$, for $\mu \geqq \mu_1$, and for every value of n; this follows from the theorem in § 454. It then follows that

$$\left|\int_b^c \cos\{(\mu+n)x-\mu a\}\lambda''(x)\,dx\right| < \epsilon, \text{ when } \mu \geqq \mu_1.$$

The terms of the first part of the series are, for $\mu \geqq \mu_1$, numerically less than the corresponding terms of the series

$$\sum_{n=1} \epsilon\frac{|a_n|}{n^2};$$

therefore the sum of the series is numerically less than

$$\epsilon\sum_{n=1}^{\infty}\frac{|a_n|}{n^2},$$

which is arbitrarily small. Therefore the sum of the first part of the series converges to zero, as μ is indefinitely increased. In a similar manner, it can be shewn that the third part of the series has the same property. Next, let us consider the second part of the series.

Choose an integer n_1, such that

$$\left|\int_b^c \cos(px-\mu a)\lambda''(x)\,dx\right| < \epsilon, \text{ for } p \geqq n_1,$$

and also such that $|a_n|<\epsilon$, for $n \geqq n_1$; it then follows that

$$\sum_{n=n_1+1}^{\infty}\frac{1}{n^2}|a_n| < \epsilon. \quad \text{We shall denote } \sum_1^{\infty}\frac{1}{n^2}|a_n| \text{ by } k.$$

Let μ_1 be any value of μ greater than $2n_1$, so that

$$\left|\int_b^c \cos\{(\mu-n)x-\mu a\}\lambda''(x)\,dx\right| < \epsilon, \text{ for } \mu \geqq \mu_1 \text{ and } n \leqq n_1.$$

We then divide the series into parts

$$\sum_{n=1}^{n_1} + \sum_{n=n_1+1}^{\iota-2} + \sum_{n=\iota-1}^{\iota+2} + \sum_{n=\iota+3}^{n_1+\iota+1} + \sum_{n=n_1+\iota+1}^{\infty},$$

and consider these five portions separately. We have, at once,

$$\left|\sum_{n=1}^{n_1}\right| < \frac{\epsilon\mu_1^2}{2(\mu_1-n_1)^2}\sum_{n=1}^{n=n_1}\frac{1}{n^2}|a_n| < \frac{k\epsilon\mu_1^2}{2(\mu_1-n_1)^2} < 2k\epsilon;\text{ for } \mu \geqq \mu_1.$$

Let the integer ι be so chosen, for any particular value of μ, that

$$\iota < \mu - 1 \leqq \iota + 1;$$

we then have

$$\left|\sum_{n=n_1+1}^{\iota-2}\right| < \frac{K\epsilon}{2\mu}\sum_{n=n_1+1}^{\iota-2}\frac{\frac{1}{\mu}}{\left(1-\frac{n}{\mu}\right)^2\left(\frac{n}{\mu}\right)^2} < \frac{K\epsilon}{2\mu}\int_{\frac{n_1}{\mu}}^{\frac{\iota-1}{\mu}}\frac{dx}{x^2(1-x)^2}$$

where K denotes $\int_c^b |\lambda''(x)|\,dx$. For, if the interval of integration be divided into portions

$$\left(\frac{n_1}{\mu}, \frac{n_1+1}{\mu}\right), \left(\frac{n_1+1}{\mu}, \frac{n_1+2}{\mu}\right)\dots,$$

we obtain the series by giving the integrand its least value, in each portion. On evaluation of the integral, we have

$$\left|\sum_{n=n_1+1}^{\iota-2}\right| < \frac{K\epsilon}{2\mu}\left(\frac{\mu}{n_1} - \frac{\mu}{\iota-1} + \frac{\mu}{\mu-\iota+1} - \frac{\mu}{\mu-n_1} + 2\log\frac{\iota-1}{n_1}\cdot\frac{\mu-n_1}{\mu-\iota+1}\right)$$
$$< \frac{1}{2}K\epsilon\left(\frac{3}{2}+\frac{4}{\mu_1}\log\mu_1\right) < K\epsilon\left(\frac{3}{4}+\frac{2}{e}\right),\text{ for } \mu \geqq \mu_1.$$

Next, we have

$$\left|\sum_{n=\iota-1}^{\iota+2}\right| < \frac{\epsilon}{2}\left\{\frac{\mu^2}{(\iota-1)^2}+\frac{\mu^2}{\iota^2}+\frac{\mu^2}{(\iota+1)^2}+\frac{\mu^2}{(\iota+2)^2}\right\}\cdot\int_b^c|\lambda(x)|\,dx$$
$$< 4\epsilon\int_b^c|\lambda(x)|\,dx.$$

An upper limit for $\left|\sum_{n=\iota+3}^{n_1+\iota+1}\right|$ may be obtained by a method precisely similar to that employed above in the case of $\left|\sum_{n=n_1+1}^{\iota-2}\right|$; it can thus be shewn to be less than a certain multiple of ϵ, for all values of $\mu \geqq \mu_1$.

It can also be shewn, as in the case of the first portion of the series, that

$$\left|\sum_{n=n_1+\iota+1}^{\infty}\right| < 2k\epsilon,\text{ for } \mu \geqq \mu_1.$$

The numbers ϵ, n_1 having been first fixed, μ_1 can then be fixed so that, for $\mu \geqq \mu_1 > 2n_1$, the sum

$$\sum_{n=1}^{\infty} -\frac{\mu^2}{2n^2(\mu-n)^2}a_n\int_b^c \cos\{(\mu-n)x-\mu a\}\lambda''(x)\,dx$$

is numerically less than a certain multiple of the arbitrarily chosen number ϵ. Therefore this part of the series converges to zero. The fourth part of the original series may be considered in the same manner; and thus the sum of the original series converges to zero, as μ is indefinitely increased.

We have now to consider the expression

$$\mu^2 \int_b^c (C + C'x + \tfrac{1}{2}A_0x^2) \cos \mu (x-a) \lambda(x)\, dx.$$

It can be verified that

$$\mu^2 (C + C'x + \tfrac{1}{2}A_0x^2) \cos \mu (x-a) = -\frac{d^2}{dx^2}\left[\left\{C - \frac{3A_0}{\mu^2} + C'x + \tfrac{1}{2}A_0x^2\right\} \cos \mu (x-a) - 2(C' + A_0x)\frac{\sin \mu (x-a)}{\mu}\right];$$

hence, on integrating twice by parts, the expression to be considered takes the form

$$\int_b^c \left[2(C' + A_0x)\frac{\sin \mu (x-a)}{\mu} \lambda''(x) - \left\{C - \frac{3A_0}{\mu^2} + C'x + \tfrac{1}{2}A_0x^2\right\} \cos \mu (x-a) \lambda''(x)\right] dx.$$

By a further application of the theorem of § 454, it is seen that μ' can be so determined that this integral is numerically less than ϵ, for $\mu \geqq \mu'$.

It has now been shewn that

$$\mu^2 \int_b^c F(x) \cos \mu (x-a) \lambda(x)\, dx$$

converges to zero, as μ is indefinitely increased, uniformly for all values of a.

483. Riemann has employed the theorems (1) and (3) of § 480, to obtain the necessary and sufficient conditions that a periodic function may be representable by means of a trigonometrical series such that the coefficients tend to the limit zero, as n is indefinitely increased. His theorem may be stated as follows:—

If $f(x)$ be a function, of period 2π, defined for every value of x, the necessary and sufficient conditions that a trigonometrical series

$$\tfrac{1}{2}a_0 + \sum_1^\infty (a_n \cos nx + b_n \sin nx),$$

such that $\lim_{n=\infty} a_n = 0$, $\lim_{n=\infty} b_n = 0$, exists, which, at every point of convergence, converges to the value $f(x)$, are the following:—

(1) *That a continuous function $F(x)$ should exist, such that*

$$\frac{F(x+\alpha+\beta) - F(x+\alpha-\beta) - F(x-\alpha+\beta) + F(x-\alpha-\beta)}{4\alpha\beta}$$

converges to $f(x)$, as α, β are diminished indefinitely in such a manner that their ratio has a finite limit.

(2) *That, if b, c be any two numbers,*

$$\mu^2 \int_b^c F(t) \cos \mu (t-x)\, \lambda(t)\, dt$$

should converge to the limit zero, as μ is indefinitely increased; where $\lambda(t)$ is a continuous function such that $\lambda'(t)$ is continuous, and that $\lambda''(t)$ is a limited function, and such that $\lambda(t)$, $\lambda'(t)$ vanish at b and c.

That the conditions are necessary has been already established in §§ 481, 482. To prove that the conditions are sufficient, let $F(t+2\pi) - F(t)$ be denoted by $\phi(t)$, then, from the condition (1), we deduce that, since $f(t+2\pi) = f(t)$,

$$\lim_{h=0} \frac{\phi(t+h) + \phi(t-h) - 2\phi(t)}{h^2} = 0.$$

Applying Schwarz's theorem, established in § 211, it follows that $\phi(t)$ must be a linear function of t. It is now seen that A_0 and C' can be so determined that $F(t) - C't - \frac{1}{2}A_0t^2$ is periodic, and of period 2π.

The condition (2) holds, not only for $F(t)$, by hypothesis, but also if $F(t)$ be replaced by $C't + \frac{1}{2}A_0t^2$, as has been proved in § 482. Denoting by $\psi(t)$ the periodic function $F(t) - C't - \frac{1}{2}A_0t^2$, it follows that

$$\lim_{\mu=\infty} \mu^2 \int_b^c \psi(t) \cos \mu(t-x)\, \lambda(t)\, dt = 0.$$

Let* $b < -\pi$, $c > \pi$; and let $\lambda(t) = 1$, in the interval $(-\pi, \pi)$; then

$$\lim_{\mu=\infty} \left[\mu^2 \int_{-\pi}^{\pi} \psi(t) \cos \mu(t-x)\, dt + \mu^2 \int_b^{-\pi} \psi(t) \cos \mu(t-x)\, \lambda(t)\, dt \right.$$
$$\left. + \mu^2 \int_{\pi}^{c} \psi(t) \cos \mu(t-x)\, \lambda(t)\, dt \right] = 0.$$

Now let μ be an integer n; we have then

$$\lim_{n=\infty} \left[n^2 \int_{-\pi}^{\pi} \psi(t) \cos n(t-x)\, dt + n^2 \int_{b+2\pi}^{c} \psi(t) \cos n(t-x)\, \lambda_1(t)\, dt \right] = 0,$$

where $\lambda_1(t) = \lambda(t-2\pi)$, in the interval $(b+2\pi, \pi)$ of t; and where $\lambda_1(t) = \lambda(t)$ in the interval (π, c). The function $\lambda_1(t)$ satisfies the conditions in (2), for the interval $(b+2\pi, c)$; hence we have

$$\lim_{n=\infty} n^2 \int_{b+2\pi}^{c} \psi(t) \cos n(t-x)\, \lambda_1(t)\, dt = 0,$$

and therefore also

$$\lim_{n=\infty} n^2 \int_{-\pi}^{\pi} \psi(t) \cos n(t-x)\, dt = 0.$$

Now let

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \psi(t)\, dt = C, \quad \frac{1}{\pi} \int_{-\pi}^{\pi} \psi(t) \cos n(t-x)\, dt = -\frac{A_n}{n^2}$$

* Weber's notes on Riemann's memoir have here been utilized.

where A_n is of the form $a_n \cos nx + b_n \sin nx$; then, in virtue of the result just established, $\lim_{n=\infty} A_n = 0$. It follows that the Fourier's series

$$C - \frac{A_1}{1^2} - \frac{A_2}{2^2} - \ldots - \frac{A_n}{n^2} - \ldots$$

is convergent for every value of x, and therefore converges to the value $\psi(x)$, this function being continuous and periodic.

The function $F(x)$ defined by the series

$$C + C'x + \tfrac{1}{2}A_0 x^2 - A_1 - \frac{A_2}{2^2} - \ldots - \frac{A_n}{n^2} - \ldots$$

satisfies, with respect to the function $f(x)$, the conditions of theorems (1) and (3) of § 480, and $\lim_{n=\infty} A_n = 0$, for every value of x. It therefore follows, from theorem (1), that the series $A_0 + A_1 + A_2 + \ldots + A_n + \ldots$, for any value of x for which it is convergent, converges to the value $f(x)$.

It will be observed that the theorem gives no information as to whether the series $A_0 + A_1 + A_2 + \ldots + A_n + \ldots$ is a Fourier's series or not; neither does it make any assertion as to the values of x for which the series is convergent.

When $f(x)$ is a given periodic function, defined for all values of x, and which satisfies the conditions of the theorem, the process of forming the trigonometrical series which represents $f(x)$ at each point at which that series converges is as follows:—The function $F(x)$ is first determined, so as to satisfy the condition (1), in relation to the given function $f(x)$, and then the periodic function $\psi(x)$ can be determined. This latter function is then replaced by the Fourier's series which represents it; and thus a series which everywhere represents $F(x)$ is obtained. The required series is then obtained by differentiating twice the terms of the series which represents $F(x)$.

Conversely, Riemann's method gives a process of summation of a given trigonometrical series $A_0 + A_1 + \ldots + A_n + \ldots$, such that $\lim_{n=\infty} A_n = 0$, for every value of x. The series being given, the function $F(x)$ is defined by the convergent series

$$C + C'x + \tfrac{1}{2}A_0 x^2 - \frac{A_1}{1^2} - \frac{A_2}{2^2} - \ldots - \frac{A_n}{n^2} - \ldots -.$$

The convergent series

$$A_0 + \sum_{n=1}^{\infty} A_n \left(\frac{\sin nh}{nh}\right)^2,$$

of which the sum is

$$\frac{F(x+2h) + F(x-2h) - 2F(x)}{4h^2},$$

is then formed. Whenever the given series is convergent, its sum is then given as the limit, when $h=0$, of the sum of the series

$$A_0 + A_1\left(\frac{\sin h}{h}\right)^2 + A_2\left(\frac{\sin 2h}{2h}\right)^2 + \dots.$$

484. The following theorem gives the necessary and sufficient condition that the trigonometrical series obtained in accordance with the method of § 483, actually converges to $f(x)$, for a particular value of x.

Let $b < x < c \leqq b + 2\pi$, and let $\rho(t)$ be a function such that $\rho(t)$, $\rho'(t)$, are continuous in the interval (b, c), and both vanish at b and at c; further let $\rho''(t)$, $\rho'''(t)$, $\rho^{\text{iv}}(t)$ be continuous in the interval (b, c). For $t = x$, let $\rho(t) = 1$, $\rho'(t) = 0$, $\rho''(t) = 0$. The necessary and sufficient condition that the series $A_0 + A_1 + A_2 + \dots$, may be convergent, in which case it necessarily converges to the value $f(x)$, is that

$$\frac{1}{2\pi}\int_b^c F(t)\,\rho(t)\,\frac{d^2}{dt^2}\frac{\sin\frac{1}{2}(2n+1)(x-t)}{\sin\frac{1}{2}(x-t)}\,dt$$

should converge to a definite limit, as n is indefinitely increased.

We have

$$A_1 + A_2 + \dots + A_n = \frac{1}{\pi}\int_a^{a+2\pi}\{F(t) - C't - \tfrac{1}{2}A_0t^2\}\sum_{r=1}^{r=n} - r^2\cos r(x-t)\,dt$$

$$= \frac{1}{2\pi}\int_a^{a+2\pi}\{F(t) - C't - \tfrac{1}{2}A_0t^2\}\,\frac{d^2}{dt^2}\frac{\sin\frac{1}{2}(2n+1)(x-t)}{\sin\frac{1}{2}(x-t)}\,dt.$$

Let a be so chosen that the interval $(a, a+2\pi)$ encloses the interval (b, c), and consider the integral

$$\int_a^{a+2\pi}\psi(t)\,\lambda(t)\,\frac{d^2}{dt^2}\frac{\sin\frac{1}{2}(2n+1)(x-t)}{\sin\frac{1}{2}(x-t)}\,dt;$$

where $\lambda(t) = 1 - \rho(t)$, when t is in (b, c), and $\lambda(t) = 1$ in the remainder of $(a, a+2\pi)$, and thus $\lambda(x) = 0$; also let $\lambda(t)$ be periodic, and of period 2π.

The integral may be expressed in the form

$$-\left(\frac{2n+1}{2}\right)^2\int_a^{a+2\pi}\psi(t)\,\lambda_1(t)\sin\tfrac{1}{2}(2n+1)(x-t)\,dt$$

$$-(2n+1)\int_a^{a+2\pi}\psi(t)\,\lambda_2(t)\cos\tfrac{1}{2}(2n+1)(x-t)\,dt$$

$$+\int_a^{a+2\pi}\psi(t)\,\lambda_3(t)\,\frac{\sin\frac{1}{2}(2n+1)(x-t)}{\sin\frac{1}{2}(x-t)}\,dt,$$

where

$$\lambda_1(t) = \lambda(t)\operatorname{cosec}\tfrac{1}{2}(x-t),$$

$$\lambda_2(t) = \lambda(t)\frac{d}{dt}\operatorname{cosec}\tfrac{1}{2}(x-t),$$

$$\lambda_3(t) = \lambda(t)\sin\tfrac{1}{2}(x-t)\frac{d^2}{dt^2}\operatorname{cosec}\tfrac{1}{2}(x-t).$$

Since $\lambda(x)=0$, $\lambda'(x)=0$, $\lambda''(x)=0$, and since $\lambda'''(t)$, $\lambda^{\text{iv}}(t)$ are continuous functions in the interval of t, it follows that $\lambda_1(t)$, $\lambda_2(t)$ satisfy the condition of the theorem (3) of § 480. Therefore the first two terms of the above expression converge to zero, when n is indefinitely increased.

The function $\psi(t)\lambda_3(t)$ has a limited differential coefficient in the interval $(\alpha, \alpha+2\pi)$, and therefore it is expressible by a Fourier's series which everywhere converges to the value of the function. The sum of the first $n+1$ terms is expressed, at the point x, by the third integral in the above expression, multiplied by $1/2\pi$; and this converges, as n is indefinitely increased, to the value of the function at the point x, that is to zero.

It has now been shewn that

$$\lim_{n=\infty}\frac{1}{2\pi}\int_\alpha^{\alpha+2\pi}\psi(t)\frac{d^2}{dt^2}\frac{\sin\frac{1}{2}(2n+1)(x-t)}{\sin\frac{1}{2}(x-t)}dt$$
$$=\lim_{n=\infty}\frac{1}{2\pi}\int_b^c\psi(t)\,\rho(t)\frac{d^2}{dt^2}\frac{\sin\frac{1}{2}(2n+1)(x-t)}{\sin\frac{1}{2}(x-t)}dt,$$

provided the limit on the right-hand side exists.

By partial integration, we have

$$\frac{1}{2\pi}\int_b^c(Ct+\tfrac{1}{2}A_0t^2)\,\rho(t)\frac{d^2}{dt^2}\frac{\sin\frac{1}{2}(2n+1)(x-t)}{\sin\frac{1}{2}(x-t)}dt$$
$$=\frac{1}{2\pi}\int_b^c\frac{d^2}{dt^2}\left[(Ct+\tfrac{1}{2}A_0t^2)\,\rho(t)\right]\frac{\sin\frac{1}{2}(2n+1)(x-t)}{\sin\frac{1}{2}(x-t)}dt.$$

The functions equal to $t^2\rho''(t)$, $t\rho''(t)$, $t\rho'(t)$, $\rho(t)$ respectively, in the interval (b, c), and each equal to zero in the remainder of the interval $(\alpha, \alpha+2\pi)$, are all representable by convergent Fourier's series, since these functions have limited derivatives in the whole interval $(\alpha, \alpha+2\pi)$. It follows that

$$\frac{1}{2\pi}\int_b^c\frac{d^2}{dt^2}\left[(Ct+\tfrac{1}{2}A_0t^2)\,\rho(t)\right]\frac{\sin\frac{1}{2}(2n+1)(x-t)}{\sin\frac{1}{2}(x-t)}dt,$$

converges to the value A_0, as n is indefinitely increased; for $\rho(x)=1$, $\rho'(x)=0$, $\rho''(x)=0$.

It has therefore been shewn that $A_0+A_1+\ldots+A_n+\ldots$ converges to a definite limit, provided

$$\frac{1}{2\pi}\int_b^c F(t)\,\rho(t)\frac{d^2}{dt^2}\frac{\sin\frac{1}{2}(2n+1)(x-t)}{\sin\frac{1}{2}(x-t)}dt$$

converges to a definite limit. The theorem has accordingly been established.

Since the interval (b, c) which contains x, is arbitrarily small, this theorem puts in evidence the fact that *the convergence, at the point x, of the trigonometrical series obtained by Riemann's process, corresponding to a given function $f(x)$, depends only upon the nature of $f(x)$ in an arbitrarily small neighbourhood of the point x.*

INVESTIGATIONS SUBSEQUENT TO THOSE OF RIEMANN.

485. The important discovery, made by Seidel and by Stokes, of the fundamental distinction between series which converge uniformly, and those which converge non-uniformly in a prescribed interval, remained for a long time without influence upon the development of the theory of series in general, and in particular of trigonometrical series. It was shewn by Weierstrass that the legitimacy of term by term integration of a convergent series is dependent upon the uniform convergence of the series; by previous writers no such restriction upon the universal validity of the process had been recognized. It was first pointed out by Heine* that a full recognition of the consequences of the theory of uniformity of convergence made it necessary to undertake a re-examination of the foundations of the theory of trigonometrical series. The investigations of Dirichlet and others had established that a function which satisfies certain conditions can be represented by means of a trigonometrical series in which the coefficients have the form given by Fourier; unless however it be assumed that a series so obtained converges uniformly, it cannot be immediately proved that it is the only trigonometrical series by which the function can be represented. The customary proof that a function is capable only of a single representation by means of a trigonometrical series was based upon the assumption that, if a convergent series $\frac{1}{2}a_0 + \sum\limits_1 (a_n \cos nx + b_n \sin nx)$ converge to zero for all values of x in the interval $(-\pi, \pi)$, it is legitimate to multiply the series by $\cos nx$ or $\sin nx$, and then to integrate term by term, between the limits $-\pi, \pi$; thus shewing that $a_n = 0$, $b_n = 0$, for every value of n. If however it is not known that the series converges uniformly, the process of term by term integration is not necessarily legitimate, and thus the proof is invalid. In fact it is conceivable that a non-uniformly convergent series might exist whose sum is zero for every value of the variable. It thus appeared that, when a Fourier's series exists which represents a function $f(x)$, it cannot be immediately inferred that no other trigonometrical series exists which represents the same function.

A Fourier's series that represents a function $f(x)$ which has discontinuities, is certainly non-uniformly convergent in the neighbourhood of such continuities, and in default of proof to the contrary, it may also be non-uniformly convergent in the neighbourhood of points at which $f(x)$ is continuous. Thus, for example, if $f(x)$ is continuous in its whole domain, and is representable by a Fourier's series, it cannot be assumed that the series is uniformly convergent. The value of the representation of a function $f(x)$ by a series $\frac{1}{2}a_0 + \Sigma (a_n \cos nx + b_n \sin nx)$ would be seriously impaired, if it were not known that the series was, at all events in general,

* *Crelle's Journal*, vol. LXXI, 1870; see also *Kugelfunctionen*, vol. I, p. 55.

uniformly convergent. For it could not be assumed that, if $\psi(x)$ denotes a continuous function, the integral $\int_a^b f(x)\psi(x)\,dx$ would be represented by the series

$$\tfrac{1}{2}a_0\int_a^b \psi(x)\,dx + \Sigma\int_a^b (a_n\cos nx + b_n\sin nx)\,\psi(x)\,dx;$$

the employment of Fourier's series in physical and other investigations would consequently be much restricted.

These considerations gave rise to a series of investigations with the view of establishing the uniqueness of the representation of a function by means of a trigonometrical series, and of investigating whether the coefficients in the series are necessarily expressible in the Fourier form. Heine* proved that the Fourier's series which represents a limited function that satisfies the conditions known as Dirichlet's, viz. that it has only a finite number of discontinuities and is in general monotone, is uniformly convergent in the portions of the interval $(-\pi, \pi)$ which remain when arbitrarily small neighbourhoods of the points of discontinuity are removed from the interval. This property of the series, of being in general uniformly convergent, suffices to remove, in the case of a most important class of functions, the restriction which has been above mentioned relating to those applications of Fourier's series which involve a term by term integration. It having thus been shewn that a function satisfying Dirichlet's conditions is representable by a series which converges in general uniformly, Heine proved that, if a function is representable at all by a series which converges in general uniformly, there can exist only one such series. This is equivalent to the theorem that, if a series converges in general uniformly in the interval $(-\pi, \pi)$, and represents zero, then all the coefficients vanish, and the sum of the series is therefore zero for all values of the variable. Heine proved further that this theorem holds even when, for a finite number of values of the variable, the series is not known to converge, or when it is at least not assumed that its sum is zero for such values of the variable. The possibility remained however, that when a function is thus uniquely represented by means of a series which is in general uniformly convergent, other series not possessing this property of uniform convergence may exist, which also represent the same function.

It was next proved by G. Cantor† that if the expression $a_n\cos nx + b_n\sin nx$ be such that, for every value of x in a given interval (α, β), the limit $\lim\limits_{n=\infty}(a_n\cos nx + b_n\sin nx)$ is zero, then a_n, b_n converge to zero, as n is indefinitely increased, and hence that the series

$$\tfrac{1}{2}a_0 + \Sigma\,(a_n\cos nx + b_n\sin nx)$$

* *Crelle's Journal*, vol. LXXI.

† *Crelle's Journal*, vol. LXXII, also in a simplified form in *Math. Annalen*, vol. IV (1871).

can only converge for all values of x in (α, β) if a_n, b_n have the limit zero, as n is increased indefinitely. This theorem is independent of any assumption that the convergence is uniform. Cantor* then deduced that, if a trigonometrical series $\frac{1}{2}a_0 + \Sigma(a_n \cos nx + b_n \sin nx)$ converges to zero, for every value of x with the exception of a finite number of values, for which it is unknown whether the series converges, all the coefficients a_n, b_n must vanish. Kronecker* shewed that this theorem can be proved without assuming the previous one. These proofs depend upon the use of Schwarz's theorem that if $F(x)$ denotes a function which is such that

$$\lim_{\epsilon=0} \frac{F(x+\epsilon) - 2F(x) + F(x-\epsilon)}{\epsilon^2} = 0,$$

then $F(x)$ must be a linear function of x.

The next step† was made by G. Cantor in extending the proof of the uniqueness of the representation of a function by means of a trigonometrical series to the case in which the function may have an indefinitely great number of points of discontinuity, these points forming a set of the first species. Starting with Weierstrass' theorem, that an infinite set of points possesses at least one limiting point, Cantor developed the theory of the successive derivatives of a set of points, and proved that if a limited function has discontinuities which form a set, one of whose derivatives contains only a finite number of points, then if the function is representable by a trigonometric series at all, there can be only one such series. In this connection the theory of sets of points was first considered, and thus the whole development of this subject, and of the more abstract theory of transfinite numbers, arose historically from the requirements of the theory of trigonometrical series. Proofs were given by Dini‡ and Ascoli§ that, for restricted classes of functions, a series which represents such functions must be Fourier's series.

An important advance in the theory was made by Du Bois Reymond‖, who proved that, if a function $f(x)$ can be represented by a series

$$\tfrac{1}{2}a_0 + \Sigma(a_n \cos nx + b_n \sin nx),$$

which is such that a_n, b_n have the limit zero, as n is indefinitely increased, the coefficients must always have the form

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,dx, \quad a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,dx, \quad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,dx,$$

whenever these expressions have a meaning. This theorem includes the

* *Crelle's Journal*, vol. LXXIII (1871).

† *Math. Annalen*, vol. V (1872).

‡ *Sopra la serie di Fourier*. Pisa, 1872, p. 247.

§ *Annali di Matematica*, ser. 2, vol. VI, p. 252, also *Math. Annalen*, vol. VI, 1873.

‖ *Abhandlungen der bayerischen Akademie*, vol. XII, 1875.

theorem as to the uniqueness of the representation of such integrable functions as are representable by series for which $\lim_{n=\infty} a_n = 0$, $\lim_{n=\infty} b_n = 0$.

The most general formulations of the theorems as to the uniqueness of the representation of a function by a trigonometrical series are due to Harnack and Hölder; an account of their results will be here given.

THE LIMITS OF THE COEFFICIENTS OF A TRIGONOMETRIC SERIES.

486. The most general form of the theorem of Cantor, that a series

$$\tfrac{1}{2}a_0 + \sum_1^\infty (a_n \cos nx + b_n \sin nx)$$

which converges for every value of x in an interval, with the exception of a certain set of points, must be such that $\lim_{n=\infty} a_n = 0$, $\lim_{n=\infty} b_n = 0$, is due to Harnack*, who proved that the theorem holds, provided that, if δ be any arbitrarily chosen positive number, a sub-interval exists in the given interval, such that, at every interior point of that sub-interval, the measure of divergence of the series is less than δ. The term "*measure of divergence of a series at a point*" is used to denote the excess of the upper, over the lower, limit of indeterminacy at the point. For each point x, at which the measure of divergence of the series is less than δ, there is a value m, of n, such that, if

$$n \geqq m, \text{ then } |a_n \cos nx + b_n \sin nx| < \delta:$$

we suppose that a sub-interval exists at every point of which this condition is satisfied. If x be any fixed point within this sub-interval, a neighbourhood $(x-\eta, x+\eta)$, of x, can be found, such that

$$|a_n \cos n(x+\epsilon) + b_n \sin n(x+\epsilon)| < \delta, \quad |a_n \cos n(x-\epsilon) + b_n \sin n(x-\epsilon)| < \delta,$$

for $n \geqq m$, provided $\epsilon < \eta$: the value of m will depend in general upon the value of ϵ. We deduce at once

$$|(a_n \cos nx + b_n \sin nx)\cos n\epsilon| < \delta, \quad |(a_n \sin nx - b_n \cos nx)\sin n\epsilon| < \delta;$$

and thence, on multiplying by $\cos nx \sin n\epsilon$, $\sin nx \cos n\epsilon$, and adding the two expressions in the inequalities, we have $|a_n \sin 2n\epsilon| < 4\delta$, and similarly $|b_n \sin 2n\epsilon| < 4\delta$, where $n \geqq m$. These inequalities hold for every δ, the values of η, m depending on the value of δ.

Let $2\epsilon = \alpha$, $4\delta = \delta'$; then for each value of α in a certain interval (a, b), n can be found, such that

$$|a_n \sin n\alpha|, \ |a_{n+1} \sin \overline{n+1}\,\alpha|, \ \ldots \ |a_{n+s} \sin (n+s)\alpha| \ldots$$

are all $< \delta'$. Suppose, if possible, that a sequence $a_{n_1}, a_{n_2}, a_{n_3} \ldots$ can be found, all of whose members are numerically $\geqq \delta''$, where $\delta'' > \delta'$; it will then be proved that a certain value of α in (a, b) can be found, such that the sequence $a_{n_1} \sin n_1\alpha$, $a_{n_2} \sin n_2\alpha$, $a_{n_3} \sin n_3\alpha \ldots$ is such that it contains an indefinitely great number of members, each of which is numerically greater

* *Bulletin des sciences mathématiques*, series 2, vol. VI, 1882, also *Math. Annalen*, vol. XIX.

than δ'. This being contrary to the hypothesis that, for each value of α, $|a_n \sin n\alpha| < \delta'$, for all values of n which are sufficiently great, leads to a contradiction; and thus it is impossible that such a sequence as $a_{n_1}, a_{n_2}, a_{n_3} \ldots$ can exist.

To establish this, we shew that, out of the sequence $a_{n_1}, a_{n_2}, a_{n_3} \ldots$, a sequence $a_{n_1'}, a_{n_2'} \ldots$ can be chosen, which is such that, for a certain value $\bar{\alpha}$, of α, in (a, b), $n_1'\bar{\alpha}, n_2'\bar{\alpha}, n_3'\bar{\alpha} \ldots$ all differ from an odd multiple of $\frac{1}{2}\pi$ by less than an arbitrarily chosen small positive number ζ.

If $$n\alpha > y_1 \frac{\pi}{2} - \zeta, \text{ and } n\alpha < y_1 \frac{\pi}{2} + \zeta,$$

then $$\frac{y_1 \frac{\pi}{2} - \zeta}{n} < \alpha < \frac{y_1 \frac{\pi}{2} + \zeta}{n};$$

now suppose n and y_1 such that

$$a < \frac{y_1 \frac{\pi}{2} - \zeta}{n}, \quad b > \frac{y_1 \frac{\pi}{2} + \zeta}{n},$$

which is equivalent to

$$\frac{2}{\pi}(na + \zeta) < y_1 < (nb - \zeta)\frac{2}{\pi}.$$

There exists a possible value of y_1 which is an odd integer, provided

$$\{n(b-a) - 2\zeta\}\frac{2}{\pi} \geqq 2, \quad \text{or if } n \geqq \frac{\pi + 2\zeta}{b - a};$$

take the least of the numbers $n_1, n_2, n_3 \ldots$ which is $\geqq \frac{\pi + 2\zeta}{b-a}$, and denote it by n_1'. We can then find a corresponding odd integer y_1, and we take α to lie in the interval (a', b'), where

$$a' = \frac{y_1 \frac{\pi}{2} - \zeta}{n_1'}, \quad b' = \frac{y_1 \frac{\pi}{2} + \zeta}{n_1'};$$

this interval lies within (a, b), and is of length $\frac{2\zeta}{n_1'}$. Next, an odd integer y_2 can be determined such that

$$(n_2'a' + \zeta)\frac{2}{\pi} < y_2 < (n_2'b' - \zeta)\frac{2}{\pi},$$

provided $$n_2' \geqq \frac{\pi + 2\zeta}{b' - a'} \geqq \frac{\pi + 2\zeta}{2\zeta} n_1';$$

and n_2' can be chosen from the sequence $n_1, n_2, \ldots$ so as to satisfy this condition, if α lies in the interval (a'', b''), where

$$a'' = \frac{y_2 \frac{\pi}{2} - \zeta}{n_2'}, \quad b'' = \frac{y_2 \frac{\pi}{2} + \zeta}{n_2'};$$

thus (a'', b'') lies within (a', b'), and is of length $\frac{2\zeta}{n_2'}$. If we proceed in this manner, we obtain a sequence n_1', n_2', ... of numbers all belonging to the sequence $n_1, n_2, n_3, \ldots$, and such that if $\bar{a}$ is the point which lies within all the intervals (a, b), (a', b'), (a'', b'') ..., then $n_1'\bar{a}$, $n_2'\bar{a}$, ... all differ by less than ζ from odd multiples of $\frac{\pi}{2}$. Since ζ can be chosen arbitrarily, we can find $n_1', n_2' \ldots$ such that $|a_{n_1'} \sin n_1'\bar{a}|$, $|a_{n_2'} \sin n_2'\bar{a}| \ldots$ are all $\geqq \delta'$, which is contrary to the hypothesis in accordance with which $|a_n \sin n\bar{a}|$ is, for every sufficiently large value of n, $< \delta'$. It has thus been shewn that no sequence $a_{n_1}, a_{n_2}, a_{n_3} \ldots$ exists, all of whose terms are numerically $\geqq \delta''$; and, if δ'' is first chosen, we may afterwards choose δ'. Therefore, from and after some value of n, a_n must be numerically less than δ'; and since this holds for every δ', we must have $\lim_{n=\infty} a_n = 0$. In a similar manner it is seen that $\lim_{n=\infty} b_n = 0$.

The theorem has thus been established that *if the series*

$$\tfrac{1}{2} a_0 + \Sigma (a_n \cos nx + b_n \sin nx)$$

be such that, for each number δ (> 0), *there exists an interval in* $(-\pi, \pi)$ *at every point of which the measure of divergence of the series is* $< \delta$, *then* a_n, b_n *diminish indefinitely, as n is indefinitely increased. In particular, if the points at which the measure of divergence is* $\geqq \delta$ *form a non-dense set, then* $\lim_{n=\infty} a_n = 0$, $\lim_{n=\infty} b_n = 0$.

It follows from this theorem that, if a trigonometrical series converges for an everywhere-dense set of points, and is such that a_n, b_n do not converge to zero, as n is indefinitely increased, then, for some positive value of δ, the set of points at which the measure of divergence is $\geqq \delta$, must be everywhere-dense. No assumption has been made as to the form of the coefficients a_n, b_n in the trigonometric series. That $\lim_{n=\infty} a_n = 0$, $\lim_{n=\infty} b_n = 0$, in the particular case of a Fourier's series, has been already established in § 454.

PROOF OF THE UNIQUENESS OF THE TRIGONOMETRICAL SERIES REPRESENTING A FUNCTION.

487. Let us assume that the series $\tfrac{1}{2} a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$ converges to zero at every point of the interval $(-\pi, \pi)$, with the exception of a reducible, and therefore enumerable, set of points, at which it is not assumed that the series converges.

In accordance with the theorem of § 486, the coefficients a_n, b_n converge to zero, as n is indefinitely increased. Accordingly the condition is satisfied that Riemann's function

$$F(x) = \tfrac{1}{4}a_0x^2 - \sum_{n=1}^{\infty} \frac{a_n \cos nx + b_n \sin nx}{n^2}$$

is a continuous function. In accordance with Riemann's theorems,

$$\lim_{h=0} \frac{F(x+h) + F(x-h) - 2F(x)}{h^2}$$

is zero, for every value of x for which the given series converges; and further, at every point, without exception,

$$\lim_{h=0} \frac{F(x+h) + F(x-h) - 2F(x)}{h} = 0.$$

In accordance with the extension of Schwarz's theorem, given in §§ 211—213, since

$$\lim_{h=0} \frac{F(x+h) + F(x-h) - 2F(x)}{h^2}$$

vanishes at every point in the interior of an interval of the everywhere-dense set of those intervals which are complementary to the given reducible set of points, and since

$$\lim_{h=0} \frac{F(x+h) + F(x-h) - 2F(x)}{h} = 0,$$

at all points, the function $F(x)$ is a linear function $ax+b$, in the whole interval $(-\pi, \pi)$, and consequently in any interval whatever.

Since the function $ax + b - \tfrac{1}{4}a_0x^2$ is represented everywhere by the series

$$-\sum_{n=1}^{\infty} \frac{a_n \cos nx + b_n \sin nx}{n^2},$$

it follows that $ax + b - \tfrac{1}{4}a_0x^2$ must be a periodic function, which can only be the case if $a = 0$, and $a_0 = 0$.

Since the series is uniformly convergent, we may multiply by $\cos nx$, or by $\sin nx$, and integrate term by term; we thus see that $a_n = 0$, $b_n = 0$. The coefficients of the series therefore vanish identically.

Since the series may be taken to be the difference of two given trigonometric series, we obtain the following theorem:—

No two distinct trigonometric series can exist, which converge to the same value for all points of $(-\pi, \pi)$, with the exception of a reducible set of points at which the series are not known to converge to the same value, or to converge at all.

This extension of Cantor's theorem, relating to the uniqueness of the representation of a function by a trigonometric series, is obtained by considering the difference of two trigonometric series which might be assumed to represent, in the sense defined, the same function. It has not been assumed that the series is necessarily a Fourier's series.

THE REPRESENTATION OF INTEGRABLE FUNCTIONS.

488. With a view to proving that, in a wide class of cases, a trigonometric series, which represents a given function, is necessarily a Fourier's series, we proceed to the consideration of an extension of Schwarz's theorem, given in §§ 211—213, due to Du Bois Reymond, and which has been otherwise proved and extended by Hölder and Lebesgue.

Let $F(x)$ be a function which is continuous in the interval $(x_1 - \alpha, x_1 + \alpha)$, and let $F(x_1 + \alpha) + F(x_1 - \alpha) - 2F(x_1)$ be denoted by $\Delta_\alpha^2 F(x_1)$. Let us further suppose that, for each value of x in the given interval, $\frac{\Delta_\epsilon^2 F(x)}{\epsilon^2}$ either converges to a fixed value $f(x)$, as ϵ is indefinitely diminished, or else has two finite limits of indeterminacy $\overline{f(x)}$, $\underline{f(x)}$. If, for any value of x, U and L are the upper and the lower limits of $\frac{\Delta_\epsilon^2 F(x)}{\epsilon^2}$, for all values of ϵ, such that $0 < \epsilon \leqq \epsilon_1$, then $\overline{f(x)}$, $\underline{f(x)}$ are the limits to which U and L converge respectively, when ϵ_1 is indefinitely diminished. We consequently assume that, in the whole interval $(x_1 - \alpha, x_1 + \alpha)$,

$$\lim_{\epsilon=0} \frac{\Delta_\epsilon^2 F(x)}{\epsilon^2} = f(x),$$

where $f(x)$ is considered to be determinate at points at which the limit is definite, and to be indeterminate between the limits $\overline{f(x)}$, $\underline{f(x)}$, at points in which the limit is indeterminate.

It will be further assumed that the upper limit of $\overline{f(x)}$ in the whole interval is finite, and equal to $\overline{U}$; it will also be assumed that the lower limit of $\underline{f(x)}$ is finite and equal to $\bar{L}$. The function $f(x)$ is therefore limited in the interval $(x_1 - \alpha, x_1 + \alpha)$, with $\overline{U}$, $\bar{L}$ for its upper and lower limits.

Let

$$\phi(x) = F(x) - F(x_1 - \alpha) - \frac{x - x_1 + \alpha}{2\alpha}[F(x_1 + \alpha) - F(x_1 - \alpha)] + \tfrac{1}{2}C(x - x_1 + \alpha)(x_1 + \alpha - x),$$

where C is a constant. We see that

$$\phi(x_1) = \tfrac{1}{2}\alpha^2\left[C - \frac{\Delta_\alpha^2 F(x_1)}{\alpha^2}\right];$$

and thus $\phi(x_1) \gtrless 0$, according as $C \gtrless \frac{\Delta_\alpha^2 F(x_1)}{\alpha^2}$. Let C be so chosen as to exceed $\frac{\Delta_\alpha^2 F(x_1)}{\alpha^2}$. Since $\phi(x)$ is continuous in the interval $(x_1 - \alpha, x_1 + \alpha)$, and vanishes at the points $x_1 - \alpha$, $x_1 + \alpha$, there must be at least one point z, in the interval, at which $\phi(x)$ has a maximum, and is positive.

We find that
$$\frac{\Delta_\epsilon^2 \phi(z)}{\epsilon^2} = \frac{\Delta_\epsilon^2 F(z)}{\epsilon^2} - C;$$
and, since $\phi(z)$ is a maximum of $\phi(x)$, $\Delta_\epsilon^2 \phi(z)$ is never positive, for all sufficiently small values of ϵ. Therefore the limits of indeterminacy of the limit of $\frac{\Delta_\epsilon^2 F(z)}{\epsilon^2} - C$, for $\epsilon = 0$, are neither of them positive; hence $\underline{f(z)} \leqq \overline{f(z)} \leqq C$.

Now $\bar{L} \leqq \underline{f(z)}$, and $\overline{U} \geqq \overline{f(z)}$; hence $\bar{L} \leqq C$, and this holds for any value of C that may be chosen, subject to the condition $C > \frac{\Delta_a^2 F(x_1)}{a^2}$.

It follows that $\bar{L} \leqq \frac{\Delta_a^2 F(x_1)}{a^2}$. In a similar manner, by choosing $C < \frac{\Delta_a^2 F(x_1)}{a^2}$, and considering the minimum of $\phi(x)$, it can be shewn that $\overline{U} \geqq \frac{\Delta_a^2 F(x_1)}{a^2}$. The following theorem* has accordingly been established:—

If $F(x)$ be continuous in the interval $(x_1 - a, x_1 + a)$, and at every point of the interval,
$$\lim_{\epsilon=0} \frac{F(x+\epsilon) + F(x-\epsilon) - 2F(x)}{\epsilon^2} = f(x),$$
where $f(x)$ is either determinate, or indeterminate between definite limits of indeterminacy, at each point x of the interval, and is a limited function in the interval, then
$$\frac{F(x_1+a) + F(x_1-a) - 2F(x_1)}{a^2}$$
lies between the upper and lower limits of $f(x)$, in the interval $(x_1 - a, x_1 + a)$.

489. Let us assume that the series
$$\tfrac{1}{2}a_0 + \sum_1^\infty (a_n \cos nx + b_n \sin nx)$$
has, throughout the interval (a, b), the sum $f(x)$, where $f(x)$ is limited in the interval (a, b), and has a determinate value at every point of the interval, except at the points of a set E of zero measure, where the values of $f(x)$ may be indeterminate, between finite limits of indeterminacy. Also let it be assumed that the conditions of the theorem of § 486, are satisfied, so that $\lim_{n=\infty} a_n$ and $\lim_{n=\infty} b_n$ are zero, or more generally, that $|a_n \cos nx + b_n \sin nx|$ is limited for all the values of n and x.

* See Hölder, *Math. Annalen*, vol. xxiv, p. 183. The theorem has also been established otherwise by Lebesgue, for the case in which $f(x)$ has a definite value at each point; see the *Annales sc. de l'école normale supérieure*, ser. 3, vol. xx.

In accordance with the theorem proved in § 481, if $F(x)$ denote the continuous function defined, for the interval (a, b), by

$$F(x) = \tfrac{1}{4} a_0 x^2 - \sum_1^\infty \frac{a_n \cos nx + b_n \sin nx}{n^2},$$

$\lim\limits_{\alpha=0} \dfrac{\Delta_\alpha^2 F(x)}{\alpha^2}$ is equal to $f(x)$, at all points of (a, b) which do not belong to $\underline{E}$; and, at a point of E, the limit, whether it be definite or not, lies between the two numbers

$$\tfrac{1}{2}\{\overline{f(x)} + \underline{f(x)}\} \pm \tfrac{1}{2}\{\overline{f(x)} - \underline{f(x)}\}\left(1 + \frac{1}{\pi} + \frac{1}{\pi^2}\right),$$

where $\overline{f(x)}$, $\underline{f(x)}$ denote the upper, and lower, limits of indeterminacy of $f(x)$ at the point.

Since $\dfrac{\Delta_\alpha^2 F(x)}{\alpha^2}$ is, in accordance with the theorem of § 488, limited in the interval, for all values of x and α, we have*

$$\int_c^{x_1} f(x)\,dx = \lim_{\alpha=0} \int_c^{x_1} \frac{F(x+\alpha) + F(x-\alpha) - 2F(x)}{\alpha^2}\,dx,$$

where c is a fixed point in (a, b), and x_1 is any point in (a, b). The integral of $f(x)$ exists, in any case, as a Lebesgue integral.

Denoting by $F_1(x)$, the indefinite integral of $F(x)$, we have

$$\int_c^{x_1} f(x)\,dx = \lim_{\alpha=0} \frac{\Delta_\alpha^2 [F_1(x_1) - F_1(c)]}{\alpha^2}.$$

Next, let $F_2(x)$ denote the indefinite integral of $F_1(x)$; we then have,

$$\int_c^x dx_1 \int_c^{x_1} f(x)\,dx = \lim_{\alpha=0} \left[\frac{\Delta_\alpha^2 F_2(x)}{\alpha^2} - \frac{\Delta_\alpha^2 F_2(c)}{\alpha^2} - (x-c)\frac{\Delta_\alpha^2 F_1(c)}{\alpha^2}\right],$$

where x is any point in (a, b). Since

$$\lim_{\alpha=0} \frac{\Delta_\alpha^2 F_2(x)}{\alpha^2} = F(x), \quad \lim_{\alpha=0} \frac{\Delta_\alpha^2 F_2(c)}{\alpha^2} = F(c),$$

we have
$$F(x) = \int_c^x dx_1 \int_c^{x_1} f(x)\,dx + Ax + B,$$

where A and B are constant for the whole interval (a, b).

Next, let the series

$$\tfrac{1}{2} a_0 + \sum_1^\infty (a_n \cos nx + b_n \sin nx)$$

be such that its sum-function $f(x)$ has indefinitely great values at, or in the neighbourhood of, points belonging to an enumerable closed set, and therefore a reducible set G, of points in $(-\pi, \pi)$. In any interval (a, b), contained in

* The extension of the theorem of § 384, given in § 472, footnote, is here employed.

the interior of an interval complementary to G, the function $f(x)$ is limited, although it may be indeterminate at points of a set E of zero measure. It will be further assumed that, either $f(x)$ has a Lebesgue integral in (a, b), or else that it has a non-uniformly convergent improper integral in accordance with Harnack's definition, or with the extension of that definition given in § 392. Also, in the latter case, it will be assumed that $f(x)\cos nx$, $f(x)\sin nx$ are also integrable in accordance with Harnack's definition, or its extension. It will now be assumed that $\lim_{n=\infty} a_n = 0$, and $\lim_{n=\infty} b_n = 0$.

Since $F(x) - \int_c^x dx_1 \int_c^{x_1} f(x)\,dx$ is linear in any interval contained in one of the intervals complementary to G, it is linear in that complementary interval, since it is continuous at the ends of that interval. Moreover c need not be interior to the interval, but may be any fixed point, say the point $x = 0$, in the interval $(-\pi, \pi)$. Therefore the continuous function

$$\psi(x) = F(x) - \int_0^x dx_1 \int_0^{x_1} f(x)\,dx$$

is linear in each interval complementary to the reducible closed set G.

We have now

$$\frac{\Delta_\epsilon^2 \psi(x)}{\epsilon} = \frac{\Delta_\epsilon^2 F(x)}{\epsilon} - \frac{1}{\epsilon}\int_x^{x+\epsilon} dx_1 \int_0^{x_1} f(x)\,dx + \frac{1}{\epsilon}\int_{x-\epsilon}^{x} dx_1 \int_0^{x_1} f(x)\,dx;$$

and, by Riemann's theorem (2), of § 480, we have $\lim_{\epsilon=0} \frac{\Delta_\epsilon^2 F(x)}{\epsilon} = 0$, at every point of $(-\pi, \pi)$ without exception. The function $\int_0^{x_1} f(x)\,dx$ being a continuous function of x_1, we see that

$$\frac{1}{\epsilon}\left\{\int_x^{x+\epsilon} dx_1 \int_0^{x_1} f(x)\,dx - \int_{x-\epsilon}^{x} dx_1 \int_0^{x_1} f(x)\,dx\right\}$$

cannot exceed, in absolute value, the fluctuation of $\int_0^{x_1} f(x)\,dx$ in the interval $(x-\epsilon, x+\epsilon)$ of x_1; and therefore the limit of the expression, for $\epsilon = 0$, is zero. It has now been proved that $\lim_{\epsilon=0} \frac{\Delta_\epsilon^2 \psi(x)}{\epsilon} = 0$, at every point of $(-\pi, \pi)$, without exception.

As in § 213, it now follows that the function $\psi(x)$ is linear in the whole interval $(-\pi, \pi)$. For, if $A_1 x + B_1$, $A_2 x + B_2$ be the values of $\psi(x)$ in the intervals which abut on one another at a point ξ belonging to G, we have $A_1\xi + B_1 = A_2\xi + B_2$; and since $\lim_{\epsilon=0} \frac{\Delta_\epsilon^2 \psi(\xi)}{\epsilon} = 0$, we find that $A_1 = A_2$, and therefore $A_1 x + B_1$ and $A_2 x + B_2$ are one and the same linear function.

The following theorem has now been established:—

If the sum-function $f(x)$ of the series

$$\tfrac{1}{2}a_0 + \sum_1^\infty (a_n \cos nx + b_n \sin nx),$$

for which $\lim_{n=\infty} a_n = 0$, $\lim_{n=\infty} b_n = 0$, be limited, or else have indefinitely great values at, or in the neighbourhood of, points in $(-\pi, \pi)$, belonging to a reducible closed set G, the function being limited in any interval contained in the interior of an interval complementary to G, and being determinate at each point of such interval, with the possible exception of the points of a set of zero measure; and if $f(x)$ either (1) *have a Lebesgue integral in $(-\pi, \pi)$, or* (2) *be such that $f(x)$, $f(x)\cos nx$, $f(x)\sin nx$ have non-absolutely convergent integrals in $(-\pi, \pi)$, in accordance with Harnack's definition, or its extension, then*

$$F(x) - \int_0^x dx_1 \int_0^{x_1} f(x)\, dx$$

is a linear function in $(-\pi, \pi)$; where $F(x)$ denotes Riemann's function

$$\tfrac{1}{4}a_0 x^2 - \sum_1^\infty \frac{1}{n^2}(a_n \cos nx + b_n \sin nx).$$

490. Let $\chi(x_1)$ denote the function $\int_0^{x_1} f(x)\, dx$, where x_1 is in the interval $(-\pi, \pi)$. The functions $F(x), f(x)$ are defined for all values of x, as periodic functions, of period 2π; if then we define $\chi(x)$ for values of x, not in the interval $(-\pi, \pi)$, by $\int_0^x f(x)\, dx$, we have

$$\chi(x_1 + 2r\pi) = \chi(x_1) + r\int_{-\pi}^{\pi} f(x)\, dx = \chi(x_1) + r\{\chi(\pi) - \chi(-\pi)\},$$

where r is some integer.

The function $F(x) - \int_0^x \chi(x)\, dx$ is continuous for all values of x; and changing x into $2r\pi + x$ only adds a constant to the function. Therefore $F(x) - \int_0^x \chi(x)\, dx$ is equal to the same linear function, for all values of x; let this linear function be $\lambda x + \mu$. The function

$$\int_0^x \chi(x)\, dx - \tfrac{1}{4}a_0 x^2 + \lambda x + \mu$$

is periodic, and of period 2π. The differential coefficient $\chi(x) - \tfrac{1}{2}a_0 x + \lambda$, of this function, is also periodic, and therefore its values for $x = \pi$, $x = -\pi$ are identical, and thus

$$a_0 = \frac{1}{\pi}\{\chi(\pi) - \chi(-\pi)\} = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\, dx.$$

Since $F(x) - \tfrac{1}{4}a_0 x^2$ is represented by the uniformly convergent series

$$-\sum_1^\infty \frac{a_n \cos nx + b_n \sin nx}{n^2},$$

we have
$$\int_{-\pi}^{\pi} \{F(x) - \tfrac{1}{4}a_0x^2\} \cos nx\,dx = -\frac{a_n}{n^2}\pi,$$
$$\int_{-\pi}^{\pi} \{F(x) - \tfrac{1}{4}a_0x^2\} \sin nx\,dx = -\frac{b_n}{n^2}\pi.$$

In these expressions, we substitute $\chi(x) + \lambda$ for $F'(x)$, after integrating by parts; we thus find that
$$\int_{-\pi}^{\pi} \{\chi(x) + \lambda - \tfrac{1}{2}a_0x\} \sin nx\,dx = \frac{a_n}{n}\pi,$$
$$\int_{-\pi}^{\pi} \{\chi(x) + \lambda - \tfrac{1}{2}a_0x\} \cos nx\,dx = -\frac{b_n}{n}\pi.$$

Let us denote the integrals
$$\int_0^x f(x) \cos nx\,dx, \quad \int_0^x f(x) \sin nx\,dx$$
by $C_n(x)$, $S_n(x)$; in accordance with our assumption, these integrals exist, either as Lebesgue integrals, in which case they are absolutely convergent, and their existence follows as a necessary consequence of the existence of $C_0(x)$, or else as non-absolutely convergent improper integrals.

In an interval (a, b) which contains no points of infinite discontinuity of $f(x)$, we have, by integration by parts, in accordance with § 394,
$$\int_a^b \{\chi(x) - \tfrac{1}{2}a_0x\} \sin nx\,dx$$
$$= \left[-\{\chi(x) - \tfrac{1}{2}a_0x\}\frac{\cos nx}{n}\right]_a^b + \frac{1}{n}\int_a^b \{f(x) - \tfrac{1}{2}a_0\} \cos nx\,dx.$$

Let now the function $\phi(x)$ be defined by means of the equation
$$\int_{-\pi}^{x} \{\chi(x) - \tfrac{1}{2}a_0x\} \sin nx\,dx = \left[-\{\chi(x) - \tfrac{1}{2}a_0x\}\frac{\cos nx}{n}\right]_{-\pi}^{x}$$
$$+ \frac{1}{n}\{C_n(x) - C_n(-\pi)\} - \frac{1}{2}\frac{a_0}{n}\int_{-\pi}^{x} \cos nx\,dx + \phi(x);$$
the function $\phi(x)$ is continuous in the interval $(-\pi, \pi)$, and it is constant in the interval (a, b). The interval (a, b) being any interval contained in a complementary interval of a reducible closed set, it follows (see § 206) that $\phi(x)$ is constant throughout (a, b); and since $\phi(-\pi) = 0$, the constant is zero. Let $x = \pi$, we then have
$$\int_{-\pi}^{\pi} \{\chi(x) - \tfrac{1}{2}a_0x\} \sin nx\,dx = \frac{1}{n}\{C_n(\pi) - C_n(-\pi)\};$$
therefore
$$a_n = \frac{1}{\pi}\{C_n(\pi) - C_n(-\pi)\} = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x) \cos nx\,dx.$$

In a similar manner, it can be shewn that
$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x) \sin nx\,dx.$$

It has therefore been shewn that, on the assumptions made, the coefficients of the trigonometrical series necessarily have the form of Fourier's coefficients.

The series is therefore either a Fourier's series, or a generalized Fourier's series, in accordance with the terminology introduced in § 439.

The following theorem has now been established:—

Let the function $f(x)$ be defined by the series

$$\tfrac{1}{2}a_0 + \sum_1^\infty (a_n \cos nx + b_n \sin nx),$$

in the sense that, at every point of convergence of the series, the sum is the value of $f(x)$, and at every point at which the sum of the series oscillates between limits, $f(x)$ is multiple-valued between those limits, and at every point of divergence of the series $f(x)$ is indefinitely great. Then, in the following cases, the coefficients of the series necessarily have the forms

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,dx,\quad a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,dx,\quad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,dx,$$

so that the series is either a Fourier's series, or else a generalized Fourier's series:—

(1) *If $f(x)$ be everywhere definite and limited. The function is then necessarily summable, and the series is a Fourier's series.*

(2) *If $f(x)$ be limited, but not everywhere definite, and satisfy the condition of the theorem of § 486, so that $\lim_{n=\infty} a_n = 0$, $\lim_{n=\infty} b_n = 0$; and further if $f(x)$ have a Lebesgue integral in $(-\pi, \pi)$, then the series is a Fourier's series.*

(3) *If $f(x)$ have infinite discontinuities at a reducible set of points, and possess a Lebesgue integral in $(-\pi, \pi)$, and the series be such that $\lim_{n=\infty} a_n = 0$, $\lim_{n=\infty} b_n = 0$, then the series is a Fourier's series.*

(4) *If $f(x)$ have infinite discontinuities at a reducible set of points, and $f(x)$, $f(x)\cos nx$, $f(x)\sin nx$ possess non-absolutely convergent improper integrals in (a, b), the series being such that $\lim_{n=\infty} a_n = 0$, $\lim_{n=\infty} b_n = 0$; then the series is a generalized Fourier's series.*

THE CONVERGENCE OF A TRIGONOMETRICAL SERIES AT A POINT.

491. If the series $\tfrac{1}{2}a_0 + \Sigma(a_n \cos nx + b_n \sin nx)$ converge at a point x, and if the sum-function of the series have definite limits $f(x+0)$, $f(x-0)$, on the right, and on the left, at the point x, it does not necessarily follow that the series converges at points in a neighbourhood of the point x, at which the series converges. From the existence of $f(x+0)$, $f(x-0)$, it follows however that, corresponding to an arbitrarily small positive number δ, a neighbourhood of the point x can be determined, such that the measure of divergence of the series is, at every point in that neighbourhood, less than δ. This has been shewn in § 486, to be a sufficient condition to ensure that the limits of a_n and b_n are zero, when n is indefinitely increased. Hence we see from Riemann's theorem, that if

$$F(x) = \tfrac{1}{4}a_0x^2 - \sum_1^\infty \frac{1}{n^2}(a_n \cos nx + b_n \sin nx),$$

then

$$\lim_{\epsilon=0} \frac{F(x+\epsilon) - 2F(x) + F(x-\epsilon)}{\epsilon^2} = f(x),$$

at the given point x, of convergence of the series.

We now have

$$2f(x) = \lim_{\epsilon=0}\left\{4\,\frac{F(x+2\epsilon)-2F(x)+F(x-2\epsilon)}{4\epsilon^2} - 2\,\frac{F(x+\epsilon)-2F(x)+F(x-\epsilon)}{\epsilon^2}\right\}$$

$$= \lim_{\epsilon=0}\left\{\frac{F(x+2\epsilon)-2F(x+\epsilon)+F(x)}{\epsilon^2} + \frac{F(x)-2F(x-\epsilon)+F(x-2\epsilon)}{\epsilon^2}\right\}.$$

In accordance with the theorem of § 488,

$$\frac{F(x+2\epsilon)-2F(x+\epsilon)+F(x)}{\epsilon^2}$$

lies between the extreme values of

$$\lim_{\alpha=0}\frac{F(z+\alpha)-2F(z)+F(z-\alpha)}{\alpha^2},$$

for $x < z < x + 2\epsilon$. It has been shewn in § 481 that, for each value of z, this limit lies between values which depend on the limits of indeterminacy of $f(z)$ at z. It follows that, for a given positive number δ, the positive number ϵ can be so determined, that

$$f(x+0) - \delta < \lim_{\alpha=0}\frac{F(z+\alpha)-2F(z)+F(z-\alpha)}{\alpha^2} < f(x+0) + \delta,$$

for every value of z, such that $x < z < x + 2\epsilon$.

We thus see that

$$\lim_{\epsilon=0}\frac{F(x+2\epsilon)-2F(x+\epsilon)+F(x)}{\epsilon^2} = f(x+0).$$

Similarly, it can be shewn that

$$\lim_{\epsilon=0}\frac{F(x-2\epsilon)-2F(x-\epsilon)+F(x)}{\epsilon^2} = f(x-0);$$

and therefore we have

$$f(x) = \tfrac{1}{2}\{f(x+0) + f(x-0)\}.$$

The following theorem has now been established:—

If a trigonometrical function converge at a point, then the value to which it converges is the mean of the limits of the sum-function, on the right, and on the left, of the point, provided those limits exist.

This theorem holds for every trigonometrical series, whether it be a Fourier's series or not.

FOURIER'S INTEGRAL REPRESENTATION OF A FUNCTION.

492. It has been shewn in the course of the investigation of conditions for the convergence of Fourier's series at a point x, that

$$\frac{1}{\pi}\int_0^{\epsilon}\frac{\sin mz}{z}f(x+2z)\,dz+\frac{1}{\pi}\int_0^{\epsilon'}\frac{\sin mz}{z}f(x-2z)\,dz$$

converges to the value

$$\tfrac{1}{2}\{f(x+0)+f(x-0)\},$$

when the positive number m, which is not necessarily integral, is indefinitely increased; provided $f(x)$ has an integral in the interval $(-\pi, \pi)$, and satisfies some one of a number of sufficient conditions in the neighbourhood of the point x, at which it is assumed that $f(x+0), f(x-0)$ exist. The numbers ϵ, ϵ' are such that

$$0<\epsilon\leqq\tfrac{1}{2}\pi,\quad 0<\epsilon'\leqq\tfrac{1}{2}\pi.$$

The above result is represented by the equality

$$\lim_{n=\infty}\frac{1}{\pi}\int_{\alpha}^{\beta}f(x')\frac{\sin u(x'-x)}{x'-x}\,dx'=\tfrac{1}{2}\{f(x+0)+f(x-0)\},$$

where $x-\pi\leqq\alpha<x<\beta\leqq x+\pi$, and where x is in the interval $(-\pi,\pi)$. If x, x' be changed into $\pi x/l, \pi x'/l$, and u be changed into ul/π, the function $f(\pi x/l)$ being replaced by $f(x)$, we see that the equality holds for points x within the interval $(-l, l)$, where α, β now satisfy the conditions

$$x-l\leqq\alpha<x<\beta\leqq x+l.$$

When $x=\alpha$, or $x=\beta$, the value of the limit is

$$\tfrac{1}{2}f(\alpha+0),\ \text{or}\ \tfrac{1}{2}f(\beta-0),$$

provided the function $f(x)$ is such that the limit exists, and also satisfies one of the sufficient conditions already referred to. For a given point x, and for given values of α, β, the number l can always be so chosen that the conditions $x-l\leqq\alpha<x<\beta\leqq x+l$ are satisfied.

Now let $f(x)$ be defined for the unlimited interval $(-\infty, \infty)$, and be not necessarily periodic, but be such that in any finite interval whatever, $f(x)$ has a Lebesgue integral. It will further be assumed that

$$\int_a^b|f(x)|\,dx$$

has a definite double limit, for $b=+\infty$, $a=-\infty$.

When $f(x)$ has a Riemann integral in (a, b), this double limit is, in accordance with the definition in § 292, denoted, whenever it exists, by

$$\int_{-\infty}^{\infty}|f(x)|\,dx.$$

We shall use this notation, even when $\int_a^b|f(x)|\,dx$ *exists only as a Lebesgue integral. Also, we shall denote by*

$$\int_{-\infty}^{\infty}f(x)\,dx,$$

the double limit, when it exists, of

$$\int_a^b f(x)\,dx, \text{ for } b=\infty,\ a=-\infty,$$

when the integral through the finite interval exists as a Lebesgue integral. This amounts to an extension of the definition given in § 292, of an improper integral through an infinite interval, to the case of Lebesgue integrals.

If $\beta' > \beta > x$, we have

$$\left|\int_\beta^{\beta'} f(x')\frac{\sin u(x'-x)}{x'-x}\,dx'\right| < \frac{1}{\beta-x}\int_\beta^\infty |f(x')|\,dx';$$

therefore, for a fixed point x, β may be chosen so great that the integral on the left-hand side is arbitrarily small. A similar remark applies to the lower limit α. We see therefore, that when $f(x')$ is such that

$$\int_{-\infty}^{\infty} |f(x')|\,dx'$$

exists, in accordance with the above definition,

$$\int_{-\infty}^{\infty} f(x')\frac{\sin u(x'-x)}{x'-x}\,dx'$$

exists, and differs from $\displaystyle\int_\alpha^\beta f(x')\frac{\sin u(x'-x)}{x'-x}\,dx,$

by less than an arbitrarily chosen positive number $\frac{1}{2}\epsilon$, provided α and β are numerically sufficiently great, for every value of u.

It follows that, when the assumed conditions are satisfied,

$$\frac{1}{\pi}\int_{-\infty}^{\infty} f(x')\,\frac{\sin u(x'-x)}{x'-x}\,dx'$$

does not differ from $\frac{1}{2}\{f(x+0)+f(x-0)\}$ by more than ϵ, provided u is equal to, or greater than, some positive number. Since ϵ is arbitrarily small, we have

$$\lim_{u=\infty}\frac{1}{\pi}\int_{-\infty}^{\infty} f(x')\frac{\sin u(x'-x)}{x'-x}\,dx' = \tfrac{1}{2}\{f(x+0)+f(x-0)\}.$$

The following theorems have now been established:—

If $f(x)$ be a function which has a Lebesgue integral in the interval (α, β), then, at a point x, in the interior of (α, β), at which $f(x+0)$, $f(x-0)$ exist, and such that one of the known conditions for the convergence of Fourier's series is satisfied by $f(x)$ in the neighbourhood of that point x,

$$\lim_{n=\infty}\frac{1}{\pi}\int_\alpha^\beta f(x')\frac{\sin u(x'-x)}{x'-x}\,dx' = \tfrac{1}{2}\{f(x+0)+f(x-0)\}.$$

In particular, this holds at every point x interior to the interval (α, β), if $f(x)$ be a function with limited total fluctuation in (α, β). At the points $x=\alpha$, $x=\beta$, the values of the limit are $\frac{1}{2}f(\alpha+0)$, $\frac{1}{2}f(\beta-0)$, provided these functional limits exist, and $f(x)$ satisfy one of the known sufficient conditions in a neighbourhood of α or β, on the side towards the interior of the interval.

If x be exterior to the interval (α, β), the limit is zero.

If $f(x)$ have a Lebesgue integral in every finite interval, and

$$\int_{-\infty}^{\infty} |f(x)|\, dx$$

have a definite value, then

$$\lim_{u=\infty} \frac{1}{\pi} \int_{-\infty}^{\infty} f(x') \frac{\sin u (x'-x)}{x'-x}\, dx' = \tfrac{1}{2}\{f(x+0) + f(x-0)\},$$

for any point x, at which $f(x+0), f(x-0)$ exist, and in the neighbourhood of which $f(x)$ satisfies one of the sufficient conditions for the convergence of Fourier's series. In particular, this result holds for all values of x, if $f(x)$ have limited total fluctuation in every finite interval, and also satisfies the condition that

$$\int_{-\infty}^{\infty} |f(x)|\, dx \text{ exists.}$$

These theorems contain Fourier's* representation of a function by means of a single integral.

493. Since $$\frac{\sin u(x'-x)}{x'-x} = \int_0^u \cos v(x'-x)\, dv,$$

the single integral $\int_{-\infty}^{\infty} f(x') \frac{\sin u(x'-x)}{x'-x}\, dx'$ may be replaced by

$$\int_{-\infty}^{\infty} dx' \int_0^u f(x') \cos v(x'-x)\, dv'.$$

Therefore the theorem in § 492 may be taken to refer to

$$\lim_{u=\infty} \frac{1}{\pi} \int_{-\infty}^{\infty} dx' \int_0^u f(x') \cos v(x'-x)\, dv'.$$

It will now be shewn that the repeated integral may be replaced by the one in which the integrations are taken in the reverse order.

Let $\psi(\alpha, \beta, v)$ denote

$$\int_\alpha^\beta f(x') \cos v(x'-x)\, dx',$$

and let $\psi(v)$ denote

$$\int_{-\infty}^{\infty} f(x') \cos v(x'-x)\, dx'.$$

On the hypothesis that $\int_{-\infty}^{\infty} |f(x')|\, dx'$ exists, as the limit of a Lebesgue integral, the integral $\psi(v)$ also exists. Since $|\psi(\alpha, \beta, v)|$ is less than some fixed positive number, for all values of α, β, and for all values of v in the interval $(0, u)$, it follows, by using the theorem of § 384, that $\int_0^u \psi(v)\, dv$ is

* See the *Théorie de la Chaleur*, Chap. IX, § 416.

the limit of the integrals $\int_0^u \psi(\alpha, \beta, v)\,dv$, when α and β have the values in sequences which diverge to $+\infty, -\infty$ respectively. Also $\int_0^u \psi(v)\,dv$ has a value which is independent of the particular sequences of α and β. For

$$\int_0^u dv \int_\beta^{\beta'} f(x') \cos v(x'-x)\,dx',$$

or

$$\int_\beta^{\beta'} f(x') \frac{\sin u(x'-x)}{x'-x}\,dx'$$

is numerically arbitrarily small, for a sufficiently great value of β, independently of the value of $\beta' (> \beta)$; and a similar remark applies to

$$\int_0^u dv \int_{\alpha'}^{\alpha} f(x') \cos v(x'-x)\,dx'.$$

We have, therefore,

$$\begin{aligned}\int_0^u dv \int_{-\infty}^{\infty} f(x') \cos v(x'-x)\,dx' &= \lim_{\beta=\infty,\,\alpha=-\infty} \int_0^u dv \int_\alpha^\beta f(x') \cos v(x'-x)\,dx' \\ &= \lim_{\beta=\infty,\,\alpha=-\infty} \int_\alpha^\beta dx' \int_0^u f(x') \cos v(x'-x)\,dv \\ &= \int_{-\infty}^{\infty} dx' \int_0^u f(x') \cos v(x'-x)\,dv.\end{aligned}$$

It follows that

$$\lim_{u=\infty} \frac{1}{\pi} \int_0^u dv \int_{-\infty}^{\infty} f(x') \cos v(x'-x)\,dx' = \lim_{u=\infty} \frac{1}{\pi} \int_{-\infty}^{\infty} dx' \int_0^u f(x') \cos v(x'-x)\,dx',$$

whenever the limit on the right-hand side exists.

Therefore
$$\lim_{u=\infty} \frac{1}{\pi} \int_0^u dv \int_{-\infty}^{\infty} f(x') \cos v(x'-x)\,dx'$$
has the value $\frac{1}{2}\{f(x+0)+f(x-0)\}$, provided the function $f(x)$ satisfies one of the sufficient conditions already referred to.

The following theorem has now been established:—

If $f(x)$ have a Lebesgue integral in every finite interval, and be such that $\int_{-\infty}^{\infty} |f(x)|\,dx$ has a definite value, as $\lim_{\beta=\infty,\,\alpha=-\infty} \int_\alpha^\beta |f(x)|\,dx$, then

$$\frac{1}{\pi} \int_0^{\infty} dv \int_{-\infty}^{\infty} f(x') \cos v(x'-x)\,dx' = \tfrac{1}{2}\{f(x+0)+f(x-0)\},$$

for any point x, at which $f(x+0)$, $f(x-0)$ exist, and in the neighbourhood of which $f(x)$ satisfies one of the sufficient conditions for the convergence of Fourier's series. In particular, this result holds for every value of x, if $f(x)$ have limited total fluctuation in every finite interval, and also satisfies the condition that $\int_{-\infty}^{\infty} |f(x)|\,dx$ exists.

It should be observed that, although

$$\int_0^\infty dv \int_{-\infty}^\infty f(x') \cos v(x'-x)\,dx'$$

is here employed to denote

$$\lim_{u=\infty} \int_0^u dv \int_{-\infty}^\infty f(x') \cos v(x'-x)\,dx',$$

in accordance with the definition of $\int_0^\infty \chi(v)\,dv$ as the limit of the Lebesgue integral $\int_0^h \chi(v)\,dv$, when h is indefinitely increased, provided the limit exists, it is not necessarily the case that

$$\int_0^\infty dv \left| \int_{-\infty}^\infty f(x') \cos v(x'-x) \right| dx'$$

exists. For the existence of $\int_0^\infty \chi(v)\,dv$ does not necessarily imply that of $\int_0^\infty |\chi(v)|\,dv$.

The above theorem contains Fourier's* representation of a function by means of a repeated integral, frequently spoken of as Fourier's double integral, although it is not a double integral in accordance with the terminology employed in this work.

Let $f(x)$ be zero for all values of x which do not lie in the finite interval (α, β), we then obtain the following result:—

If $f(x)$ have a Lebesgue integral in (α, β), then

$$\frac{1}{\pi}\int_0^\infty dv \int_\alpha^\beta f(x') \cos v(x'-x)\,dx'$$

has the value $\frac{1}{2}\{f(x+0)+f(x-0)\}$

at an interior point of the interval (α, β), and has the values

$$\tfrac{1}{2}f(\alpha+0), \quad \tfrac{1}{2}f(\beta-0),$$

at the points α, β; provided, in each case, the limit exists, and $f(x)$ satisfies one of the requisite conditions in the neighbourhood of the point. If x be exterior to (α, β), the repeated integral is zero.

It should be observed that the repeated integral

$$\int_\alpha^\beta dx' \int_0^\infty f(x') \cos v(x'-x)\,dv$$

does not exist, because $\int_0^\infty \cos v(x'-x)\,dv$

has no definite meaning.

There is no difficulty in obtaining sufficient conditions for the uniform convergence of Fourier's repeated integral to the value of the function, in an interval in which the function is continuous, as in the case of Fourier's series.

* See the *Théorie de la Chaleur*, Chap. IX, § 1.

APPENDIX.

ON TRANSFINITE NUMBERS AND ORDER-TYPES.

A brief reference will be made to some criticisms and remarks* which have been made relating to the views expressed in the general discussion of the theory of transfinite numbers and order-types, contained in §§ 152—163. Some other writings on the subject, which have recently appeared, and were not mentioned in Chapter III, will also be noticed.

The term "norm" as used in the definition of an aggregate (§ 153) has been treated by Russell as synonymous with the term "propositional function" which he himself employs. This proceeding has led to some misconception as to the scope of the definition in § 153. It was intended that a "norm" should always be of such a character as to leave no doubt as to the existence of elements in the aggregate which it defines, whereas the corresponding implication is apparently not made when the term "propositional function" is employed. Thus it is incorrect to assert, as does Russell (*loc. cit.* p. 40), that the existence of certain classes is denied, although such classes possess unimpeachable norms. On the contrary, the existence of the classes in question was denied, on account of the absence of the necessary norms, in the sense in which the term "norm" is employed. It may be the case that an unimpeachable propositional function is present, and yet that there may be no corresponding class. No opinion is here expressed as to the most suitable terminology in the general logic of classes.

That aggregates exist which have no cardinal number has been maintained in § 157, and is also in accordance with the view of Jourdain (*loc. cit.* p. 266). The aggregate W, which contains all the ordinal numbers, is the first example of such an aggregate. To this aggregate the term "inconsistent" has been applied by Cantor and Jourdain, but it has been proposed later to abandon this adjective as descriptive of the aggregate W, on the cogent ground that there is no inconsistency in the recognition of the existence of the aggregate, but only in the attribution to it of a cardinal number, and of an order-type. The difference of view as regards the aggregate W, held by Jourdain from that here maintained, relates not to the existence, but to the constitution of W. In accordance with

* These criticisms and remarks refer to the article published in the *Proc. Lond. Math. Soc.*, ser. 2, vol. III, p. 170, which is substantially identical with §§ 152—163. They are due to Russell, *Proc. Lond. Math. Soc.*, ser. 2, vol. IV, p. 29; to Hardy, in the same volume, p. 10; and to Jourdain, in the same volume, p. 266.

the ideas expressed in § 157, so far as our present knowledge reaches, every segment of W is enumerable; whereas, in Jourdain's view, there exist segments of which the cardinal numbers are Aleph-numbers of unending variety of order.

As soon as general agreement is attained as to the fundamentals of this subject, it will be possible to introduce a more precise and suitable terminology than is at present possible. It will probably be desirable to restrict the term "aggregate" to such collections of objects as possess a cardinal number, and if they are ordered, also an order-type. Some other term would then be applied to such cases as the class W, when no cardinal number, or order-type exists. A complete theory must contain, in the first place, precise definitions of the terms "cardinal number" and "order-type," and would then set forth the necessary and sufficient conditions to be satisfied by a collection, that it might possess a cardinal number, and if ordered, then also an order-type, in accordance with the definitions of these terms.

A definition of cardinal number has been given in § 155, and it has there been pointed out, that an aggregate possesses a cardinal number, only when it is one of a plurality of equivalent aggregates essentially distinct from one another. It has been suggested by Russell (*loc. cit.* p. 40), and also by Jourdain (*loc. cit.* p. 273), that if we have one series A, we can always obtain another series similar to A, by interchanging two terms of A, or by replacing a term of A by something else, or, in the case of a normally ordered aggregate, by removing some of the terms at the beginning; and thus that the criterion, given in § 155, is in all cases satisfied.

The elements of the new aggregates so obtained would however not be essentially distinct from the original one, and the existence of such new aggregates is stated in § 155 to be insufficient as a reason for attributing the existence of an order-type, or of a cardinal number, to the original aggregate. In order to leave no doubt upon this point, it is desirable to render the statement of the conditions for the existence of a cardinal number, or of an order-type, more precise, by postulating that the given aggregate M must be one of a plurality of equivalent, or of similar, aggregates, such that the elements m' of one of these M', are not only essentially distinct and different from the elements m, of M, but are also such that the existence of these elements m' cannot be deduced as a direct consequence of the existence of the elements m of M. The finite numbers perform the function of counting objects of all sorts and descriptions; the analogous function of transfinite ordinals or cardinals, would be that of counting sets of objects which are at least not confined to belong to a very special class of objects, viz. transfinite numbers and order-types.

The criticisms contained in §§ 159—162, of those proofs of theorems which depend upon the assumption that an infinite number of arbitrary acts of choice is a valid process, have been discussed by Russell, Hardy, and Jourdain, in relation to principles known as the "axiom of Zermelo," and the "multiplicative axiom." Russell has signified (*loc. cit.* p. 47) his complete agreement with the views expressed in §§ 159—162; but Hardy and Jourdain maintain the validity of the "multiplicative axiom," in accordance with which, when an infinite number of aggregates have been defined, there exists a new aggregate, each element of which consists of an aggregate of elements one of which belongs to each of the given aggregates. It is held that this principle is valid, although it may be impossible to assign rules by which one element is to be chosen out of each of the given aggregates, and when it is consequently impossible to define any single element of the so-called multiplicative class. This principle is maintained by Hardy (*loc. cit.* pp. 14—17), and by Jourdain (pp. 281—282), as a valid postulate, partly by an appeal to authority, and on account of its utility in much interesting mathematics, and partly because it is assumed to be so self-evident that a denial of it is paradoxical. It seems however difficult to assign a precise meaning to the existence of an aggregate, when not even a single element of it is capable of real definition. Whenever the multiplicative class is, in any particular case, capable of definition, the axiom is unnecessary. The existence of an abstract mathematical object would appear to be entirely dependent on a precise definition of such object; in default of such definition, such hypothetical object is not identifiable. Thus the hypothetical elements of a multiplicative class, whenever the use of the axiom is necessary, are not identifiable as individuals, and are indistinguishable from one another.

The axiom employed by Zermelo in his attempt to prove that every aggregate is capable of being normally ordered, and which consists of the assumption that each and every part of a given aggregate can be correlated with a single element contained in that part, has been discussed* by Hadamard, Lebesgue, Baire, and Borel, in connection with Zermelo's proof (§ 161). It is there pointed out by Hadamard, that the real question at issue is whether it is possible to demonstrate the existence of mathematical entities which cannot be precisely defined; a question which Hadamard himself answers in the affirmative, but with which answer the other writers do not appear to be in agreement. The tentative character of some investigations in which use of an infinite number of acts of choice has been made is pointed out by Borel.

À propos of a criticism† by Levi, of Bernstein's proof (see § 150), that the aggregate of closed sets of points in the n-dimensional continuum has the power c of the continuum, Bernstein‡ has endeavoured to avoid the difficulty involved in the use of a correspondence which cannot be defined, by introducing the conception of "multiple equivalence." Thus, if there are two aggregates M and N, for which an aggregate $\Phi = \{\phi\}$ of reversible (1, 1) correspondences ϕ exists, in which no element is special (*ausgezeichnet*), then the two aggregates are said to be multiply equivalent. The cardinal number $\overline{\overline{\Phi}} = f$ is then termed the multiplicity of the correspondence ϕ. In case the multiplicity is unity, the aggregates are said to have a one-valued equivalence. The difficulty of this conception is the same as the one referred to above, in the case of the multiplicative axiom, viz., that the aggregate Φ is such that no single element of it is capable of definition, and that the elements are consequently indistinguishable from one another.

* "Cinq lettres sur la théorie des ensembles," *Bulletin de la Soc. Math. de France*, 1905.

† "Intorno alla teoria degli aggregati," *Lomb. Ist. Rend.* (2), 1902, pp. 863—869.

‡ "Bemerkung zur Mengenlehre," *Nachrichten der Ges. d. Wissensch. zu Göttingen*, 1904, p. 557.

The difficulties which have arisen in connection with the aggregate W, of all ordinal numbers, on the assumption that it possesses segments with transfinite cardinal numbers of infinite variety, have been dealt with* by Bernstein. He postulates that the ordinal numbers of the series W have the following properties:—(1) they are order-types of normally ordered aggregates; (2) if a be one of them, there always exists a next greater one, $a+1$; and thus a is always the order-type of segments of normally ordered aggregates. He then defines W to be the aggregate of all order-types of the segments of normally ordered aggregates, and deduces that W is itself normally ordered. He next assumes that, although W satisfies the condition (1), it does not satisfy the condition (2); and thus that W is itself not a segment of a normally ordered aggregate. This assumption as to the nature of W, which appears as a postulate, amounts to the assertion that no aggregate (W, e) can exist, in which e has higher rank than all the elements of W. That the validity of such postulation is at least doubtful, appears clearly, if we consider that the hypothetical object W is incapable of being substituted for the element A in an aggregate (A, e) in which e is regarded as of ordinally higher rank than A. It would appear that, if e is of higher rank than A, it might also be of higher rank than every element contained in A, when A is composite, or in any aggregate which may be substituted for A.

The failure of attempts to define a set of points of the continuum which shall have a (1, 1) correspondence with all the ordinal numbers of the first and second classes is connected with the fact that there exists no systematic representation of all the numbers of the second class. The usual symbolism ω, ϵ, &c. breaks down at certain points, where new symbols have to be employed for the representation of higher numbers of the class. It is however easily seen that no enumerable set of new symbols will suffice for a mode of representation of all the numbers of the class. For only an enumerable set of numbers could be represented by means of an enumerable set of symbols $\omega, \epsilon, \ldots$, all possible forms obtained by use of these symbols being employed.

An attempt has been made by König† to distinguish between those elements of the arithmetic continuum which are "finitely defined," and those which are not capable of finite definition. It can however be shewn‡ that the distinction introduced by König is an invalid one. A full discussion of this matter would require a larger amount of space than can here be given to it; reference must therefore be made to the literature in which the point is discussed.

* "Ueber die Reihe der transfiniten Ordnungszahlen," *Math. Annalen*, vol. LX, p. 187. An article by Schönflies "Ueber die logischen Paradoxien der Mengenlehre," *Jahresber. d. Deutsch. Math. Ver.*, vol. XV, 1906, may be referred to; also an article by Harward, *Phil. Mag.* Oct. 1905, p. 457.

† "Ueber die Grundlagen der Mengenlehre und das Kontinuumproblem," *Math. Annalen*, vol. LXI, 1905; also *Acta Math.*, vol. XXX, p. 329. See also an article by Dixon "On a question in the theory of aggregates," *Proc. Lond. Math. Soc.*, ser. 2, vol. IV. The same idea has been discussed in the shape of a paradox, by Richard, *Acta Math.*, vol. XXX, p. 295. For a further development of König's ideas, see *Math. Annalen*, vol. LXIII, p. 217.

‡ See Hobson, "On the Arithmetic continuum," *Proc. Lond. Math. Soc.*, ser. 2, vol. IV, where the matter is fully discussed. A reply by Dixon, "On well ordered aggregates," is given in the same volume.

LIST OF AUTHORS QUOTED.

[*The numbers refer to pages.*]

Abel, 496, 497, 501
Ampère, 265
Archimedes, 40
Arzelà, 122, 426, 474, 481, 494, 537, 540, 543
Ascoli, 52, 63, 339, 745

Baire, 65, 93, 114, 146, 237, 239, 327, 525, 530, 532, 533, 765
Baker, 653
Beman, 9
Bendixson, 57, 90, 141, 541, 562
Bernoulli (D.), 637, 639
Bernstein, 156, 159, 185, 189, 765, 766
Bessel, 641
Bettazzi, 51, 246
Birkhoff, 478
Bôcher, 482, 649, 719, 727
Bolzano, 58
Bonnet, 359, 666
Borel, 65, 86, 90, 102, 106, 122, 128, 159, 210, 228, 522, 524, 533, 707, 728, 765
Brodén, 97, 116, 117, 217, 219, 249, 285, 296, 298, 620, 695
Bromwich, 466, 589, 592
Burali-Forti, 195, 201
Burkhardt, 635, 637
Burmann, 21

Cantor (G.), 9, 21, 30, 47 *et seq.*, 62 *et seq.*, 71, 74, 76, 79, 82, 94 *et seq.*, 123, 148, 158, 163, 170, 181 *et seq.*, 194, 198 *et seq.*, 206, 211, 275, 330, 480, 618, 744, 745
Cantor (M.), 19, 35
Cathcart, 316, 606
Cauchy, 37, 222, 339, 365, 463, 500, 504, 510, 641
Césaro, 707
Couturat, 9

D'Alembert, 636
Dantscher, 323, 515
Darboux, 339, 477, 481, 620
Dedekind, 2, 20, 21, 25, 151
Dini, 235, 243, 266, 274, 285, 294, 298, 343, 349, 366, 378, 461, 471, 478, 564, 607, 620, 626, 666, 694, 701, 745
Dircksen, 641
Dirichlet, 216, 231, 247, 337, 497, 641, 664
Dixon, 766
Du Bois Reymond, 9, 36, 52, 57, 62, 64, 90, 207, 265, 290, 345, 359, 360, 366, 378, 408, 422, 430, 455, 477, 510, 585, 592, 626, 652, 701, 732, 745

Euclid, 19
Euler, 47, 636, 638

Fatou, 701, 719, 729
Fejér, 708
Forsyth, 719
Fourier, 57, 215, 639, 649, 760, 761
Frege, 1

Genocchi, 302, 321, 730
Gibbs, 653
Gibson, 637
Goursat, 87
Gravé, 275

Hadamard, 765
Hahn, 251, 351
Hamilton, 10
Hankel, 20, 98, 241, 243, 247, 361, 607, 626
Hardy, 187, 191, 207, 210, 444, 445, 462, 479, 497, 590, 600, 607, 763
Harnack, 57, 90, 98, 101, 109, 244, 248, 316, 367, 426, 725, 746
Harward, 766
Heine, 9, 21, 29, 222, 229, 672, 743, 744
Helmholtz, 2, 9, 10
Hermite, 68
Hilbert, 38, 195, 333, 335
Hobson, 131, 194, 342, 396, 442, 456, 474, 483, 486, 490, 494, 537, 540, 566, 586, 683, 702, 713, 763, 766
Hölder, 51, 53, 360, 388, 695, 751

Huntington, 169
Hurwitz, 715
Husserl, 10

Jacob, 20
Jordan, 65, 107, 136 *et seq.*, 256, 367, 417, 426, 432, 442, 468, 583, 666, 701
Jourdain, 171, 183, 195, 209, 763

Kneser, 649
König, 284, 766
Köpcke, 296, 298, 299, 626
Kowalewski, 360
Kronecker, 21, 360, 695, 698

Lagrange, 504, 637
Lambert, 47
Lebesgue, 102, 226, 259, 295, 332, 342, 390, 522, 525, 551, 554, 556, 576, 580, 646, 674, 708, 715, 751, 765
Legendre, 47
Leibnitz, 1
Lerch, 522, 620, 727
Levi, 765
Liapounoff, 716
Lindemann, 68
Liouville, 68, 97
Lipschitz, 694, 701
Locke, 1
London, 464
Loria, 329
Love, 653

Méray, 21
Mertens, 500
Meyer, 360
Michelson, 653
Mill, 1
Mittag-Leffler, 64, 81, 522
Moore, 335, 381, 384

Netto, 329, 360
Neumann (C.), 222, 263, 360, 666
Newton, 19

Osgood, 113, 480, 482, 485, 486, 540, 543

Pasch, 98, 235
Peacock, 20
Peano, 9, 23, 107, 248, 319, 323, 330
Pereno, 299, 626
Picard, 522, 719, 722
Pierpont, 426, 448
Pincherle, 21
Poincaré, 653
Poisson, 640
Pompeiu, 620

Pringsheim, 9, 19, 47, 145, 266, 269, 305, 360, 405, 426, 429, 464, 501, 505, 512

Reiff, 477
Richard, 766
Riemann, 247, 338, 341, 405, 637, 679, 730
Roche, 504
Rudio, 47
Russell, 9, 23, 30, 169, 195, 204, 763

Sachs, 637, 676, 701
Scheeffer, 265, 273, 275, 290, 323, 513, 515, 521
Schepp, 55
Schläfli, 652
Schlömilch, 504, 729
Schönflies, 90, 116, 135, 139, 205, 248, 251, 284, 299, 335, 351, 370, 379, 417, 423, 428, 484, 586, 766
Schröder, 156
Schubert, 9
Schwarz, 277, 319, 676, 701, 719
Seidel, 477
Smith (H. J. S.), 64, 97, 416
Stäckel, 672
Steinitz, 299
Stekloff, 716
Stéphanos, 48
Stokes, 458, 477, 482, 727
Stolz, 52, 54, 98, 269, 316, 319, 321, 323, 372, 381, 417, 430, 432, 442, 515, 520, 585, 592
Study, 252, 256

Tannery, 31, 480
Thomae, 302, 314, 339, 416, 428
Toepler, 724

Vallée-Poussin (de la), 368, 379, 440, 566, 571, 576, 590, 603, 715, 716
Veblen, 169
Veltmann, 145
Veronese, 51, 55
Volterra, 247, 357, 481, 522, 723

Wallis (John), 35
Weber, 739
Weierstrass, 21, 58, 224, 264, 359, 360, 477, 522, 625
Whitehead, 153
Wiener, 625

Young (W. H.), 84, 90, 102, 111, 113, 118 *et seq.*, 128, 185, 186, 342, 345, 395, 487, 540, 541

Zermelo, 156, 209

GENERAL INDEX.

[The numbers refer to pages.]

Abel's theorem on power-series, 497

Adherence, 125

Aggregate; General notion of, 1; Elements of, 1; Finite, 3; Similar, 4, 162, 200; Equivalent, 6, 150; Rank of elements in, 6; Ordinal number of, 7; Definition of finite, 8; Perfect, 49; Straight line as, 56; Power of, 70; Definition of, 149, 197; Comparable and Incomparable, 152; Closed, 166; Dense in itself, 166; Everywhere-dense, 166; General theory of, 194; Normally ordered, 170; Parts of, 151; Limiting elements of, 165; of order-types η, θ, π, 167; Segments of normally ordered, 172; Simply ordered, 163, 198; Structure of simply ordered, 165

Aleph-numbers, 154; General theory of, 181, 209

Archimedean system, 40

Archimedes, Axiom of, 54

Arithmetization, 13; Kronecker's scheme of, 21

Baire's, theorem on representation of functions by series of continuous functions, 525; classification of functions, 532

Cardinal numbers, Definition of, 8, 150, 199, 204; Addition and Multiplication of, 152; as exponents, 153; Division of, 159; exceeded by other cardinal numbers, 158; of continuum, 183, 187; of second class of ordinals, 180; Relative order of, 151; Smallest transfinite, 154; of aggregate of continuous functions, 227

Categories, First and Second, 114

Coherence, 125

Condensation of singularities, 607; Cantor's method of, 618

Congruency, Axiom of, 53

Connexity of aggregate of real numbers, 49

Content, of sets of points, 77; of closed sets, 108

Continuity, of functions, 221; Uniform, 229; of sum-function, 475

Continuous functions, 225; two definitions not equivalent, 226; defined at points of a set, 226; Oscillating, 281, 294; Construction of, 296; with respect to each of two variables, 324; Non-differentiable, 620

Continuum, of real numbers, 49; given by intuition, 52; Straight line as, 53; Arithmetic, 72; Cardinal number of, 183, 187; of n-dimensions, 190; Order-type of, 169

Convergence, General principle of, 36; of series, defined, 453; Absolute and conditional, 457; Absolute, of double series, 468; Uniform, 470; Simply-uniform, 471; Non-uniform, 474; Tests of uniform, 477; Measure of non-uniform, 484; Distribution of points of non-uniform, 485; of power-series, 495; of product of two series, 500

Correspondence, General notion of, 2

Curves filling space, 330

Derivatives, of sets of points, 61, 98; of functions, 263; Upper and lower, 265; Progressive and regressive, 265; Limited, 283; General properties of, 285
Differential coefficients of functions, 263; Successive, 275; Partial, 311; Higher partial, 316
Differentiation, of series, 561; of integrals with respect to parameters, 599; of Fourier's series, 725
Dirichlet's, integral, 656, 695; investigation of Fourier's series, 658
Discontinuity, Infinite, 220; of functions, 233; of functions, classification, 234; of first and second kinds, 235
Discontinuous functions, Classification of, 240; Point-wise, 243; Construction of point-wise, 246
Divergence, of series, 455; Measure of, 746
Double integrals, 416, 440; Properties of, 430; Improper, 432, 438; over infinite domains, 442; Transformation of, 445

Enumerable aggregate, definition, 66; rational numbers, 68; algebraical numbers, 68; isolated aggregates, 83
Equivalence theorem, 119
Extent, Exterior and Interior, 108

Fluctuation, of function, 220; Inner, 220
Fourier's series; definition, 640, 646; Formal expression of, 641; cosine and sine series, 643; Generalized, 646; Dirichlet's investigation of, 658; Dirichlet's conditions for convergence of, 664; Uniform convergence of, 669, 683; Limiting values of coefficients in, 671; Convergence of coefficients in, 675; Sufficient conditions of convergence of, 680; Non-convergence of, 701; Series of arithmetic means related to, 707; Properties of coefficients of, 715; Integration of, 718; Differentiation of, 725; Limits of coefficients in, 746
Fractional numbers, 12
Frontier of set of points, 108, 136
Functions, Dirichlet's definition, 216; Homonomically and heteronomically defined, 217; Upper and lower limits of, 219, 224; Limited, definition, 219; Unlimited, 219; Upper limit of, in an interval, 220; Continuity of, 221; continuous in an interval, 222; Limits of, at a point, 230; Maxima and minima of, 234, 259; Semi-continuous, 237; Ordinary, 299; Classification of discontinuous, 240; Point-wise discontinuous, 243; Monotone, 245, 263; defined by extension, 249; Most nearly continuous, 251; with limited total fluctuation, 252; with limited total variation, 257; with limited derivatives, 283; with one derivative assigned, 294; of two or more variables, 299; of two variables, 301; Double and Repeated limits of, of two variables, 303, 304; Existence and equality of repeated limits of, 306; Uniform convergence of, 309; Maxima and minima of, 309; linear in each interval of a set, 357; Summable, 391; represented by series, 469; Maxima and minima of, of one variable, 512, of two variables, 515; Representation of, by series of continuous functions, 522; Baire's classification of, 532; Construction of continuous non-differentiable, 620; Differentiable, everywhere-oscillating, 626
Functional image, 223
Functional limits; Aggregate of, 233
Functional relation, 214

Generation, Cantor's principle of, 76
Graphs, Functions defined by, 215

Heine-Borel theorem, 86

Incrementary ratios, 286; Limit of, 290
Indeterminate forms, Evaluation of, 268
Infinitesimals, Non-existence of, 40; Veronese's theory of, 55
Integers, Sets of sequences of, 146
Integrable functions, Particular cases of, 343; null-functions, 347
Integral numbers, Operations on, 10

Integral Calculus, Fundamental Theorem of, 349, 355, 387, 550
Integral, Definite, of limited functions, 338; Riemann's definition of, 338; Upper and lower, 339; Conditions of existence of, 341; Properties of definite, 344; Indefinite, 350; Improper, 364; Harnack's definition of improper, 367; de la Vallée-Poussin's definition of improper, 368; Absolutely and conditionally convergent, 369, 371, 377, 395; Lebesgue's definition of, 390; Equivalence of definitions of, 396; with infinite limits, 398; Change of variable in an, 410; Double, definition of, 416; Conditions for existence of double, 418; Repeated, 421; Improper double, 432, 438, 440; Non-absolutely convergent improper Lebesgue, 537; Repeated improper, 566; Repeated Lebesgue, 576; Repeated, of unlimited function, 582; Restricted Jordan double, 583; Repeated, over infinite domain, 586; Limit of, containing a parameter, 594; Poisson's, 640, 719; Dirichlet's, 656, 695; Fourier's, 758; Fourier's single, 760; Fourier's repeated, 762
Integration, Geometrical interpretation of, 388; Double, 415; Lebesgue's theory of, 390; by parts, 407, 559; of series, 584; of limit of a sequence, 538; with respect to a parameter, 582; of Fourier's series, 718
Intervals, Open and closed, 50; Unenumerable sets of, 89
Irrational numbers, 19; Development of theory of, 21; Dedekind's theory of, 18; Cantor's theory of, 26

Limits, Arithmetical theory of, 35; Method of, 36
Lines of invariability, 260

Magnitude, Theory of, 19, 52
Maxima and Minima, 234, 259; Proper and improper, 266; of functions of one variable, 512; of functions of two variables, 515
Mean value theorem, of Differential Calculus, 267; of Integral Calculus, 358; Application of, to Fourier's series, 666
Measure, of sets of points, 81, 107, 143; Exterior and interior, 108, 142; of limits of sequences of closed sets, 113; of discontinuity defined, 233
Measurement, 51

Negative numbers, 51
Null-function, Point-wise discontinuous, 350

Order, Notion of, 2
Order-functions, 185
Order-types, 168; Addition and multiplication of, 163; of continuum, 169, 206; η, θ, π, 167
Ordinal-numbers, Definition of, 4, 175; Transfinite, 74; of second class, 76, 177; Arithmetic of, 183; Limiting, 177; Sum and product of, 176
Oscillation, of function at a point, 235; of function in an interval, 236, 261

Power-series, 495; Interval of convergence of, 496; Abel's theorem on, 497

Rational numbers, Aggregate of, 18; Section of aggregate of, 23
Real numbers, Dedekind's definition, 23; Cantor's definition, 29; Operations on, 31; Convergent sequences of, 33; Equivalence of two definitions of, 38; Representation of, 45; everywhere-dense, 49
Remainder-function, Transformed, 483
Repeated, limits of functions, 304; Existence and equality of, limits of functions, 306; integrals, 421; improper integrals, 566; Lebesgue integrals, 576; integrals of unlimited functions, 582; integrals over infinite domains, 586; Fourier's, integral, 762
Riemann's theory of trigonometrical series, 730

Saltus, 233, 300; -function, 238
Schwarz's theorem, 277; on partial differential coefficients, 319
Semi-continuous functions, 237

Sequences, Simple, 6; Convergent, 26; Ascending and descending, 164; Functions defined by, 458; Double, 464; Repeated limits of double, 465; Monotone, 465; Conditions of equality of repeated limits of, 466; Integration of limits of, 538
Series, Limiting sum of, 453; Remainder in a, 453; Convergent, 453; Non-convergent arithmetic, 455; Divergent, 455; Oscillating, 455; Limits of indeterminacy of, 455; Absolutely and conditionally convergent, 457; of transfinite type, 461; Double, 463, 464, 467; Absolutely convergent double, 468; Uniform convergence of, 470; Simply uniform convergence of, 471; Taylor's, 501, 514; of continuous functions, 522; Integration of, 534; Differentiation of, 561; Special cases of trigonometrical, 638, 648; Fourier's series, 640, 646; Generalized Fourier's, 646; Formal expression of Fourier's, 646; of arithmetic means related to Fourier's, 715; Properties of coefficients of Fourier's, 715; Special cases of trigonometrical, 638; Finite trigonometrical, 722; Riemann's theory of trigonometrical, 730; Uniqueness of Fourier's, 748; Convergent trigonometrical, 756
Sets of intervals; Limiting point of, 60; Properties of, 81; External, internal, and semi-external points of, 84; Unenumerable sets of, 89
Sets of points, 57; Analysis of, 123; Closed, 64, 84; Common points of system of, 92; Connex, 138; Degrees of points in, 123; dense-in-themselves, 64; Derivatives of, 61, 98; Detached, 137; Everywhere-dense, 65; Exterior points of, 107; Frontier of, 108, 136; Improper limiting points of, 61; Inner limiting, 127; Interior points of, 107; Irreducible, 98; Isolated, 64; Limiting points of, 60; Measurable, 104; Non-dense, 65; Non-dense closed, 90; Non-dense perfect, 90, 94; of first species, 62, 83; of second species, 63; of first and second categories, 144; Perfect, 65; Reducible, 98; Separated, 65; Sequences of closed, 110, 117; Sets of closed, 93; Transfinite derivatives of, 79; Unbounded, 59; Upper and lower boundaries of, 58; of cardinal number of continuum, 191, 210
Square, Representation of, in linear interval, 329
Sum-function, Continuity of, 475, 489; Transformed, 483; Limits of a, 487

Taylor's series, 501; for functions of two variables, 514
Total fluctuation, 252
Total variation, 257

Unenumerable; Proof that continuum is, 69
Unity, Notion of, 1

Variable, Real, 213
Vibrating strings, 636

CAMBRIDGE: PRINTED BY JOHN CLAY, M.A. AT THE UNIVERSITY PRESS.